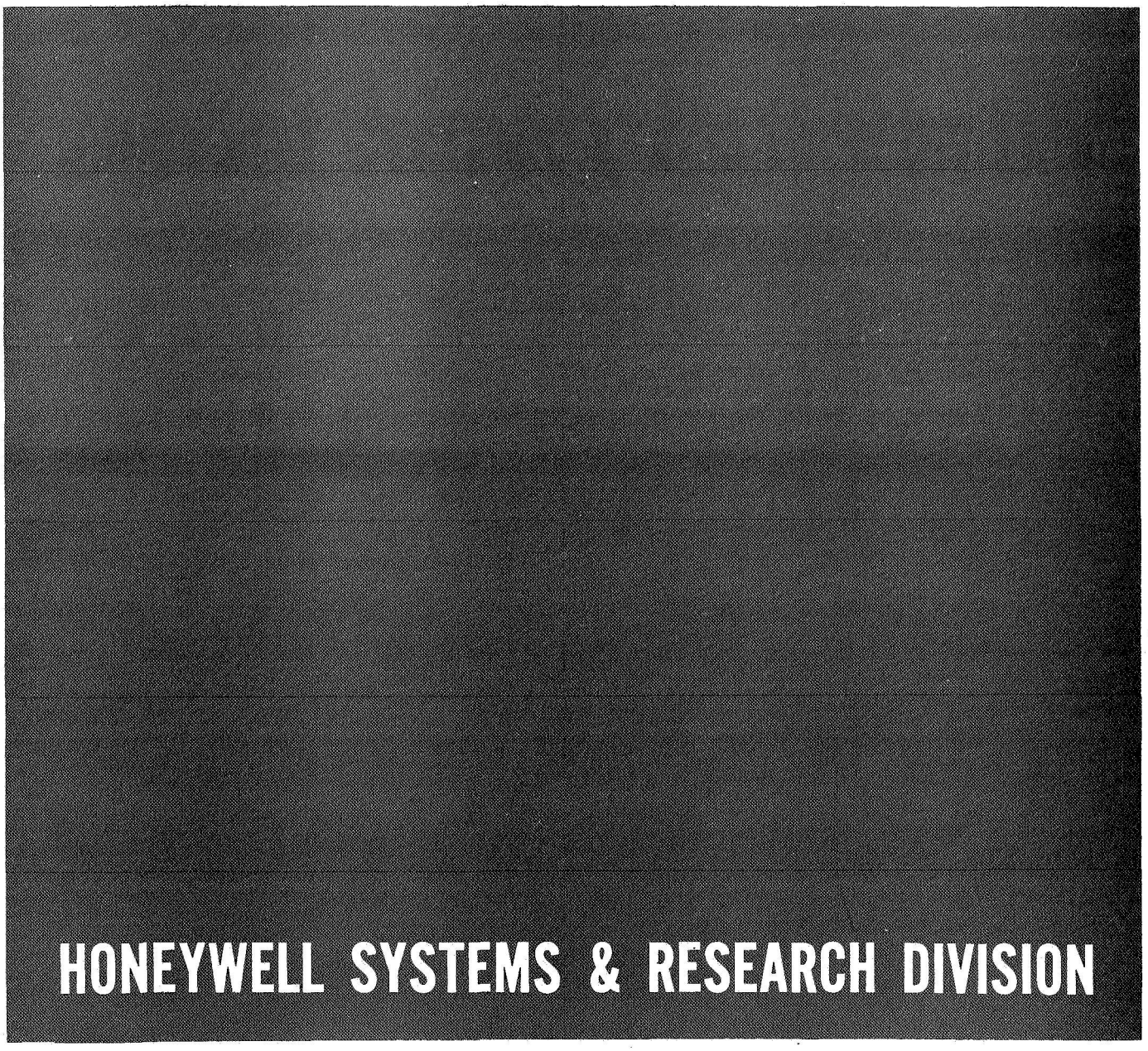
AMT
File copy

NAS-12-38
N69-35653

APPENDICES for 12017IRI



HONEYWELL SYSTEMS & RESEARCH DIVISION

APPENDIX A

COMPUTATIONAL REQUIREMENTS OF ADVANCED PLANETARY EXPLORATION

CONTENTS

APPENDIX A
COMPUTATIONAL REQUIREMENTS OF ADVANCED PLANETARY EXPLORATION

INTRODUCTION

This appendix presents a discussion of the computation requirements for an unmanned vehicle involved in planetary exploration in the period 1975-1985.

Due to a time limitation, only those functions or tasks which appear to have extensive computational requirements were selected. The five functions or tasks examined are:

- Data Handling Function
- Processing of Scientific Data
- Navigation of the Spacecraft
- Navigation and Control of Lander
- Communications and Command Processing

DATA HANDLING FUNCTION

The data handling function is divided into the following three categories - data acquisition, data storage and retrieval, and data distribution. The function is concerned with both scientific and performance data.

Data Acquisition

The data acquisition function includes selection and sampling of inputs containing both scientific and performance data, as well as conversion functions to make the data suitable for storage in a digital memory or for the communications subsystem.

Adaptive Activation of Scientific Instruments

A subset of the scientific instruments will be active for sampling at any given time. It is desired that the decision to activate an instrument be based not only on programmable things such as phase of the trip, but also on the results from presently active instruments. For example, the detection of particles of a particular type in space might call for the activation of other instruments and thus make their outputs candidates for sampling. Of course the decision to run a given experiment will also require activation of a subset of instruments.

This control is desired on an instrument level (rather than groups of instruments). It seems unlikely that more than 25 instruments would be involved.

Adaptive Selection of Performance Inputs

The subset of performance inputs to be selected for sampling will also vary during the course of the mission. This variation may depend on the phase of the trip, on the operations being performed by a given subsystem, or on the status of a subsystem or part of a subsystem.

Adaptive Sampling of Scientific Instruments

Of those instruments which are active at a given time, it is desired that each be sampled at a rate determined by considerations such as the following:

1) On the amount of change since the last reading.

2) On the relative change compared to the changes encountered by other instruments.

3) On the rate at which data can be stored.

4) On the current data rate of the telemetry system.

5) On the basis of priority.

6) When the quantity passes through a peak or a valley.

Adaptive Sampling of Performance Data

Depending on the status of the system it may be desirable to alter the sampling rate on performance data. If a subsystem is found to be operating out of tolerance in some way, sampling rates may be increased. Another situation where changes in sampling rate will occur is if there is interruption of the spacecraft to earth communications link. It is desirable to reduce the sampling rate to reduce the storage requirement.

System Status Monitoring

The purpose of at least part of the performance data is to check the status of the equipment. Thus it is desired to compare each reading with a set of limits to determine whether that subsystem is operating properly. These

checks must be made in order to be able to make an immediate decision and also to reduce the load on the communications link.

Dynamic Range Adjustment

It is necessary to have the capability of changing the range of a given input quantity in the course of the mission, since generally, the range of the instrument is much larger than the allotted word length. For instance a magnetometer with a dynamic range of 0.1 - 1000 may be allotted a word length of eight bits. Non linear encoding is sometimes used in such situations. Another solution is to consider the quantity a floating point number made up of a fraction and an exponent. The exponent is included in the sample only if it has changed since the last sample.

Conversion from Analog to Digital

Many of the performance measurements as well as a number of the scientific measurements provide analog quantities which must be converted to digital form before either storage or transmission. The significance of such data is 10 bits or less.

Storage and Retrieval Requirements

This function includes not only the requirement for a storage device but also includes a number of storage related control functions such as storage allocation and storage buffering.

Data Storage

The storage requirements for voyager have been discussed in Reference 8 for both the lander and orbiter vehicles. On the lander, the requirements are divided into video data storage, non-video scientific data storage, and

performance data storage. Video data storage is the largest requirement (7×10^8 bits) and probably is a separate unit. The storage requirement for performance data reaches its highest point during entry when the total is 14×10^6 bits. For the orbiter the scientific data storage requirement is estimated to be 6×10^8 bits and the performance data storage requirement is 700K bits. There is also a requirement to store data relayed from the lander and this is estimated at 3×10^7 bits.

One of the more difficult requirements is that of retrieving previously stored data during the same period that new data is being stored. An example of a situation where this occurs is shortly after landing when it is desired to complete transmission of entry and descent data and at the same time store results of surface experiments. Another requirement is to have the capability of reading out selected quantities from a file. This is necessary so that critical quantities from data collected during entry can be transmitted during descent. This is to guard against total loss of information in case the landing is unsuccessful.

Probably the most difficult requirement to meet is that of reliability. This is particularly true if a centralized storage unit is used.

Storage Allocation

The allocation or assignment of storage to the various types of data is necessary in some situations. One such requirement would occur if a buffer was used to accumulate data from many instruments in order that some processing operation can be performed on blocks of data. The allocation task would be that of assigning blocks of storage locations to each instrument.

Another possible requirement for storage allocation occurs if stored reference data is to be brought from a bulk storage device to a random access

device to be shared by several types of reference data at the same time, the space must be allocated depending upon the total demand. This allocation cannot be pre-programmed since the requirements will depend upon the situation encountered.

Data Retrieval

There are a number of requirements for information retrieval of a more complex nature than, for instance, simple sequential reading of all the data on a tape. Some of these are:

1) During entry of a lander vehicle a large amount of data is gathered but none can be transmitted because of the expected complete communications black out. Thus this data must all be stored. It is desirable to retrieve the most vital of this information for communications during descent in case the landing is unsuccessful. It can be seen that this also is difficult to pre-program since the duration of the communications black-out is not known.

2) Many of the processing operations require comparison of gathered data versus stored data. This stored data must be retrieved from a bulk storage device and stored where it is more readily accessible if it is planned to use it many times, and of course it must also be retrived from the fast device each time it is to be used.

3) There is likely to be the need for retrieval of all samples of a single quantity from a bulk device for data compression and transmission.

Data Distribution

There is considerable switching and selection required to get the newly acquired data to the proper storage unit or communication channel, and

also to select data from storage units and direct inputs to the communications channel. The output multiplexer is essentially a many-to-one switch. The control of devices such as the output multiplexer is done partially by stored program and partially by the contents of data being received. It thus has the same type of adaptive selection requirement as the data acquisition devices.

PROCESSING OF SCIENTIFIC DATA

The on-board data processing performed on scientific data is divided into two categories. The first is that required for data compression which is necessary to make best use of the communications capacity. The second is that required for on-board decision making.

Data Compression

The necessity of data compression techniques becomes obvious by considering the vast amount of data to be collected and the severe limitations in communicating the data to earth. In addition, efficient on-board data compression techniques will help alleviate serious problems in areas such as power and reliability.

Probably the most rewarding area for applying data compression procedures is in the processing of pictures obtained by TV or radar. However, nearly all those situations where a sequence of reading from a given instrument is taken are subject to these procedures.

A general definition of data compression is "the transformation of one ordered set of integers into another set of integers, either ordered or unordered, from which the desired information can be obtained, with the restriction that the second set be representable with fewer bits than the first. Data compression methods can be grouped into two classes:

Encoding methods are those having a unique inverse so that the original data set can be reproduced from the compressed set.

Filtering methods are those that do not possess a unique inverse.

Encoding Methods

Δ Modulation

The coded sequence is the first differences of the original sequence. In a sequence with high local correlation, but having a large overall range of values, the first differences will have a much smaller range of values and thus each element can be represented with fewer bits. For very "smooth" sequences higher differences can be used but the process becomes "noisy" as soon as adjacent coded values of opposite sign appear.

Debiasing -- If the range of values is small but the individual values are large the subtraction of the smallest value will reduce the number of bits for each element.

Interval Suppression -- If a large number of adjacent elements of the sequence have the same value then a gain will result by using the common value and the number of elements. This might be applied effectively to pictures. If the intensity within a certain block is equal or nearly equal, the average intensity and identification of the block can be sent.

Code Substitution -- The original values are replaced by new values so that the resulting sequence has some desired property. For example, in a sequence of integers from 1 to 10, suppose the values 4 and 7 never appear. Then reassigning $1 \rightarrow 0$, $2 \rightarrow 1$, $3 \rightarrow 2$, $5 \rightarrow 3$, $6 \rightarrow 4$, $8 \rightarrow 5$, $9 \rightarrow 6$, $10 \rightarrow 7$ would allow using a 3 bit code in place of a 4 bit code. If, for a particular sequence or subsequence, a small set of values occurs with high frequency then it would be worthwhile to assign these low numerical values.

As an example consider the following sequence of digits taken from a random number table

6 7 9 0 6 5 9 1 7 7 7 2 0 9 9 3 0 3 4 9 6 9 7 3 6 2 7 6 2 9 1 1 4 2 9

the frequency of occurrence of the various digits is:

N	0	1	2	3	4	5	6	7	8	9
f	3	3	4	3	2	1	5	6	0	8

To code the sequence in binary it is necessary to use 4 bits for each digit to accommodate the range 0 - 9. This requires 140 bits. If the following value substitution is made

old	0	1	2	3	4	5	6	7	8	9
new	4	5	3	6	7	8	2	1	9	0

then 32 of the 35 digits lie in the range 0 - 6 which can be represented with 3 bits per integer, but a 3 bit code cannot be used for the entire sequence. Since there are so few exceptions to the 3 bit codable portion an efficient code can be generated for these exceptions.

Let the digits 0 - 6 be coded by their 3 bit binary representations and use 1110 for 7, 1111 for 8. The 3 bit symbol 111 does not appear in the coding of the digits 0 - 6 and thus serves to identify the code shift. With reassignment and code shifting the sequence can be represented by 108 bits. This count does not include the number of bits required to describe the encoding scheme; for a short sequence this is a significant number of bits but for a long sequence it would not be prohibitive.

Filtering Methods

Filtering was defined as an approach to data compression wherein a unique inverse does not exist. In general these methods require more extensive computation than the encoding methods just described. Three methods are

described. The first has the characteristic that the order in which the data was collected is of no importance; only the value is important. The second and third methods retain the order of the sequence and realize compression by making approximations on the values.

Statistical Representation -- One method of describing the behavior of a variable is to generate the distribution function which can then be represented for transmission in several different ways.

Quantiles -- One of these methods of approximating or describing the distribution function is to generate a set of quantities called quantiles.[10] Given a cumulative distribution function F(x), which rises steadily from the value O to the value 1, this function can be specified by giving the values of the independent variable x corresponding to conveniently chosen values of the function. These values of x are called quantiles.

This definition is made clear by considering the cumulative distribution function F(x) and the probability density function f(x) (see Figure A1 of a block of 1,000 data points.) These data points might represent 1,000 successive readings of a particle count instrument where the range of the reading is 0-63.

Four values of F(x) —q_1, q_2, q_3, and q_4 — are specified. The result desired for each of these values is the corresponding value of x. It can be seen from the diagram in Figure A1 that the value of x can be obtained by generating a histogram of the probability density function and then performing a summation of counts starting at x = 0 and working toward x = 63. As the total equals or exceeds each of the specified q values, the value of x is recorded.

The values of x are then transmitted to earth to represent the 1,000 samples.

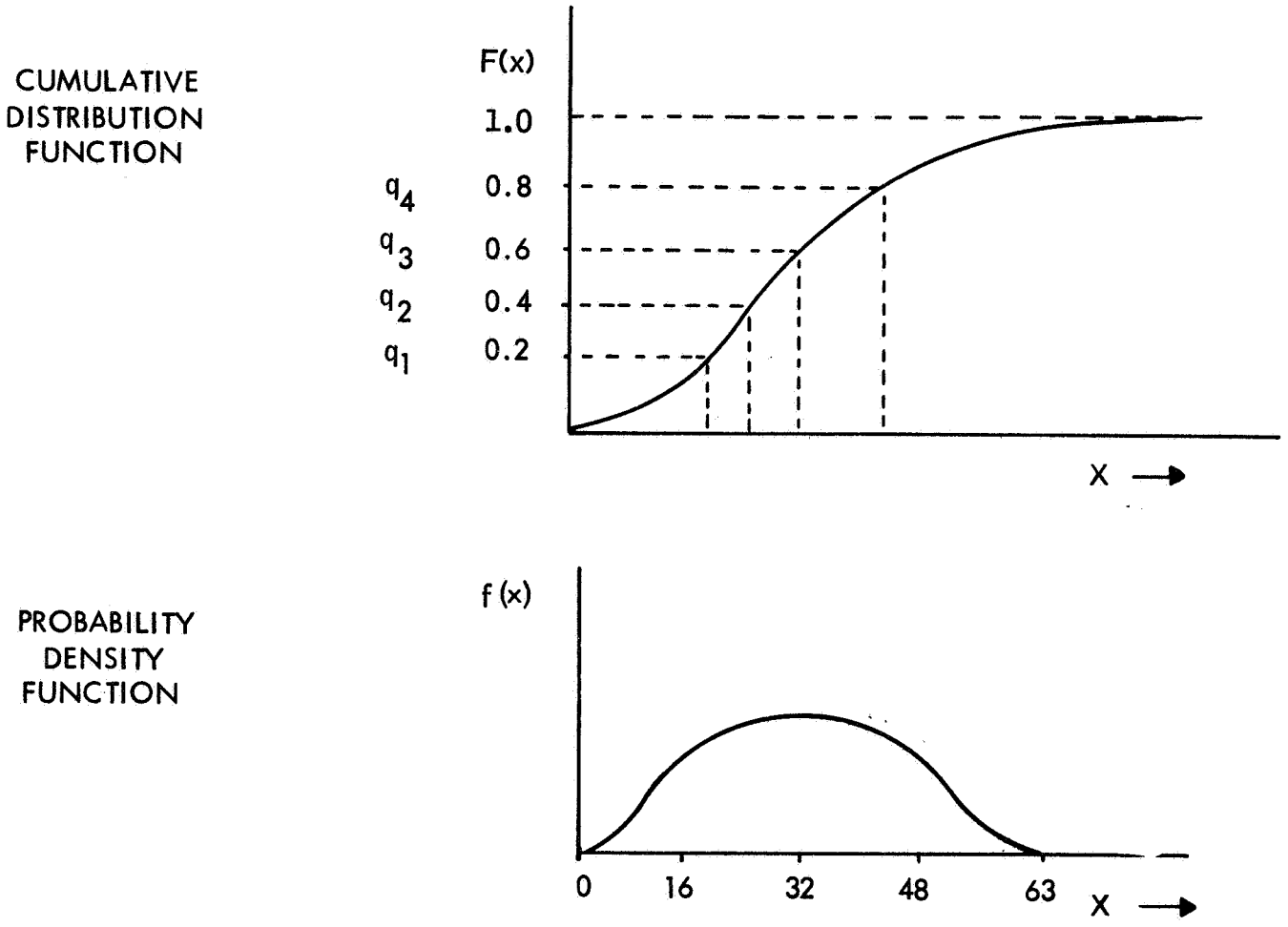


Figure A1. Four Values of F(x) -- q_1, q_2, q_3, and q_4

The main advantage of quantiles is that they can be computed very easily and because they are a set of useful descriptors of the data. In Reference 10 it is shown that the mean and standard deviation can be obtained quite accurate from quantiles. Moreover the methods are relatively insensitive to deviations from the normality.

Direct Computation of Moments -- Another approach to obtaining statistical parameters is the direct computation of moments. The first moment, the mean, is simply

$$M_1 = \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

The second moment, the variance

$$M_2 = \sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

and in general the p^{th} moment

$$M_p = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^p$$

While such an approach requires considerably more computation than the quantile method, it is more flexible and will provide a more complete description of the probability density function.

Direct Approximation of the Probability Density Function -- Another approach that is feasible since the required type of computation is similar to that required for other functions is to fit a curve to the probability density function. This would be done by generating a histogram and using a curve fitting procedure such as the least square method.

Curve Fitting -- One of the most effective methods of filtering is to use polynomial predictors or polynomial interpolators.[11]

A polynomial predictor of zero order simply predicts that the next sample will be within certain limits of the last reading and if it is the sample is not taken. Only when a new value falls outside these limits is it considered a sample. When a new sample is obtained a new set of limits is computed and the procedure is continued. It can be recognized that the effect is the same as obtained by adaptive sampling described in the Data Handling Function paragraph. Thus we will turn our attention here to polynomial interpolators.

Polynomial interpolation is a process of curve fitting to the samples after they are taken, rather than rejecting samples as in the prediction procedure. A zero-order interpolator is much like a zero order predictor, the main difference being in the value of the sample used to represent a redundant data set. The reference sample for the interpolator is computed at the end of a redundant set of points by

$$Y_r = \frac{Y_h + Y_1}{2}$$

where

Y_r = reference sample

Y_h = highest sample

Y_1 = lowest sample

Thus the reference sample is simply the average of the highest and the lowest sample in the redundant set.

A first-order polynomial interpolator approximates the values of successive samples with line segments. It is desired to find a line segment so that as many data points as possible lie within a certain distance of this line. The line then serves as a substitute for the data points.

The optimum first order algorithm requires freedom in placement of both ends of the line segment. Both the starting point and end point of each line segment are computed values. Also the end point of one line segment can be connected by a straight line to the starting point of the next.

The computation of the line segments described above is sufficiently difficult that less optimal approaches are considered. One such is called the joined line segment approach. In this case the start of the second segment is the same value as the computed end of the first. In another approach called the disjoined line segment method, the starting point of the next line segment is the first actual value excluded from the preceding line segment.

Polynomial predictors of higher order should also be considered. In this case a polynomial evaluation would be required for each point to determine if it is within a certain tolerance.

One of the problems encountered in using these data compression techniques is in setting the tolerance level. During active periods a fairly wide tolerance must be used in order to obtain reasonable compression. However if the same tolerance is used during inactive periods a very noticeable degradation is seen. To overcome this problem adaptive tolerance control (the capability of altering the tolerance on the basis of the activity level of the data) is needed.

It appears that data compression hardware should have a great deal of flexibility. This is true because the best method or technique depends to a great extent on the data input.

A curve fitting technique which might find application in a number of ways is least square curve fitting. One application of this technique is to generate the coefficients of a polynomial predictor.

The procedure for weighted least squares fitting to tabulated functions is as follows:

Let x_i, $i = 1, - - - -, n$, each denote a vector whose components are the value of a function $x_i(t)$ at the points $t_1, t_2, - - - t_m$.

Let $(x_i, x_j) = \sum_{k=1}^{m} W_k x_i(t_k) x_j(t_k)$ and $||x||^2 = (x, x)$.

The problem is to find coefficients γ_i such that for a given function (vector) f, the quantity $||f - \sum_{i=1}^{n} \gamma_i x_i||^2$ is minimum. A solution to the problem is to construct and solve the linear equations $\sum_{i=1}^{n} \gamma_i (x_i, x_k) = (f, x_k)$ where $k = 1, - - - -, n$

Thus the computation requires a sum of products operation and a matrix inversion.

Complex predictors might also take on the form of the following equation:

$$f(k+1) = a_0 f(k) + a_1 f(k-1) + - - - - - + a_{n-1} f(k-n)$$

This is called linear regression prediction. Determination of the coefficients would require a matrix inversion which could be as large as 20 x 20. Once the coefficients have been determined the predicted next value of f is obtained by a simple formula evaluation.

Correlation -- The information is extracted from the sequence by detection of similarities between it and other sequences or similarities among the subsequences. The usual example is a noise filter where the autocorrelation function of the noise is used to separate noise from signal.

Correlation procedures, in general, can be performed as a sum of products calculation.

Feature Oriented Techniques -- This class of filtering is concerned with detection and/or identification of features within a data sequence. This is primarily aimed at compression of tv or radar pictures.

The operator approach to pattern recognition is one example of a feature oriented filtering technique. In this case, since recognition is not the aim, the process would be carried only through the generation of a characteristic vector phase. If the problem were recognition, the generated characteristic vector would be compared against stored reference vectors. The generation of the characteristic vector is done by scanning the input picture by a set of operators and generating a component of the vector for each operator. Thus with k operators the characteristic vector is

$$X = (X_1, X_2, - - - - -, X_k)$$

The generation of each of these components can be done in a variety of ways but generally the type of computation required is the sum of products type.

On-Board Decision Making

The processing operations described here include an analysis of the contents of the data rather than simply reducing the data as in the previous section. Normally such procedures will be used because the results of the analysis are required on-board. It can be recognized, however, that if the results of the analysis are relayed to earth in place of the original data, the effect is the same as if a data reduction procedure had been used.

The filtering techniques described for data compression are also applicable for decision making operations. Thus in this section decision making situations rather than computational procedures are described.

Composition Analysis

Composition analysis is performed by both orbiter and lander vehicles using a variety of instruments. In some situations the results of the first part of a composition analysis experiment are required before the second part can proceed. The instrument will take complete spectra, the contents of the data will be analyzed, and then, depending on the contents, the instrument will be commanded to look at specific portions of the spectra.

Example of the type computation required for composition analysis are provided by considering first alpha scattering experiments such as would be used for surface soil analysis, and then experiments using various kinds of spectrometers as proposed for atmospheric analysis by a lander during entry and descent.

Alpha Scattering Experiment -- Alpha particles from a radio-active source bombard a sample of the soil. Some of the particles scattered from nuclei within the sample strike the surface of a small semiconductor detector placed at a high scattering angle. The amplified pulses from the detector are proportional to the energies of the incident alpha particles. These numbers are analyzed electronically and converted to digital representations. These digital numbers form the energy spectrum of the scattered particles. The mass numbers of nuclei within the target sample as well as the abundances of these nuclei can be determined from this energy spectrum. Figure A2 shows typical energy spectra obtained from mono-elemental targets. Figure A3 shows that of a multi-component sample. In Figure A3 each element is identified by the position of the breakpoint, and the abundance of each is determined by the distances between the plateaus.

The experimental procedure for a multi-component sample is a follows:

1) A series of reference spectra is stored.

2) A spectrum of the multi-component sample is generated.

3) A comparison operation is performed to determine the best match between stored spectra and various portions of the multi-component spectrum. Reference 17 suggests a least-square criterion would be used for this operation. There are obviously many other procedures that could be used as well. For instance fitting line segments to the spectrum by interpolation techniques previously described and then checking for large values of slope can be used to detect breakpoints, which identify the elements.

A correlation requirement has also been suggested to compare the compositional analysis results from one instrument with those of another. This might be useful to check and calibrate an instrument, for instance. It might also be done simply to increase the probability of drawing a correct conclusion as to the composition of the sample being tested.

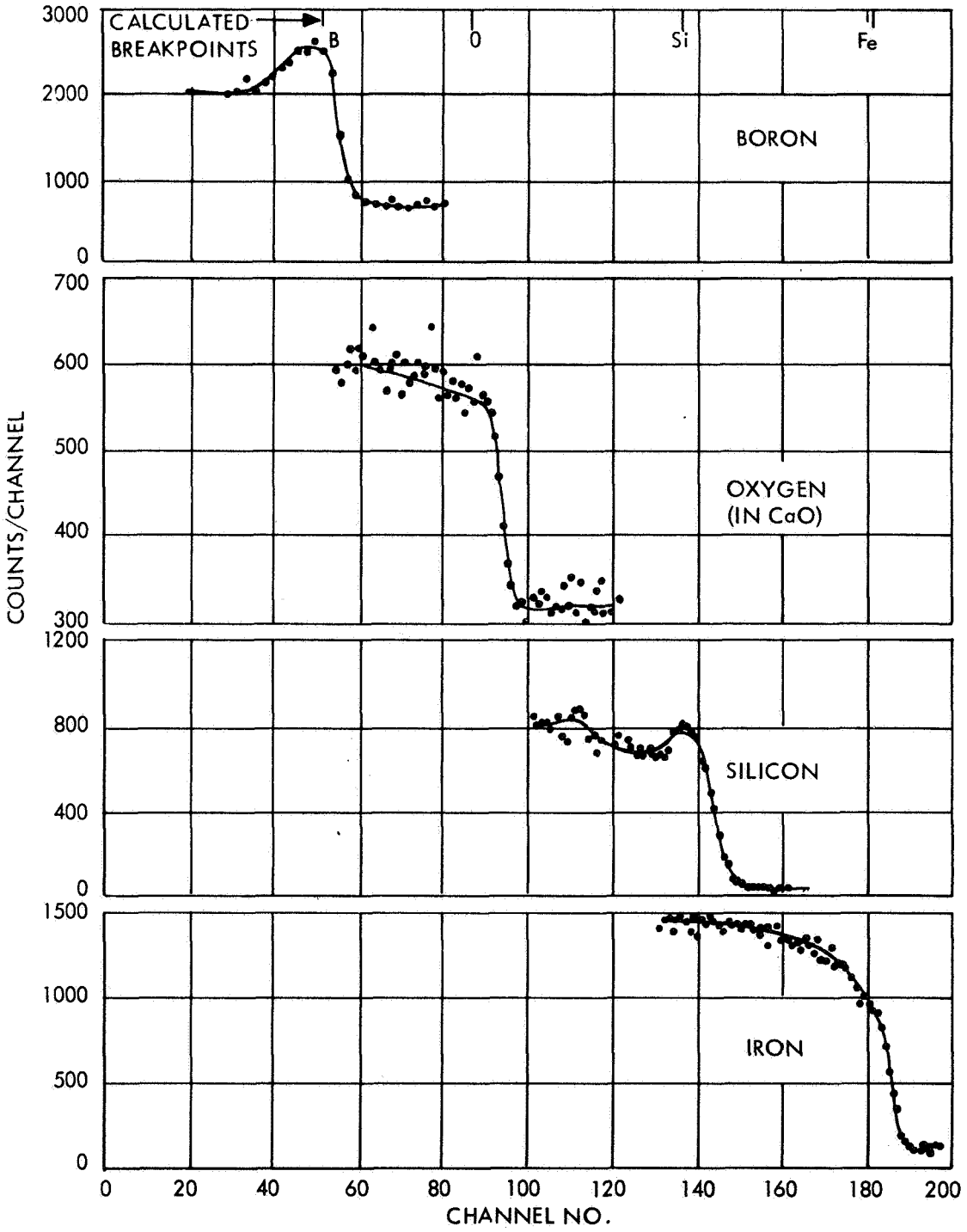


Figure A2. Alpha Scattering Spectra of Several Elements (Reference 17)

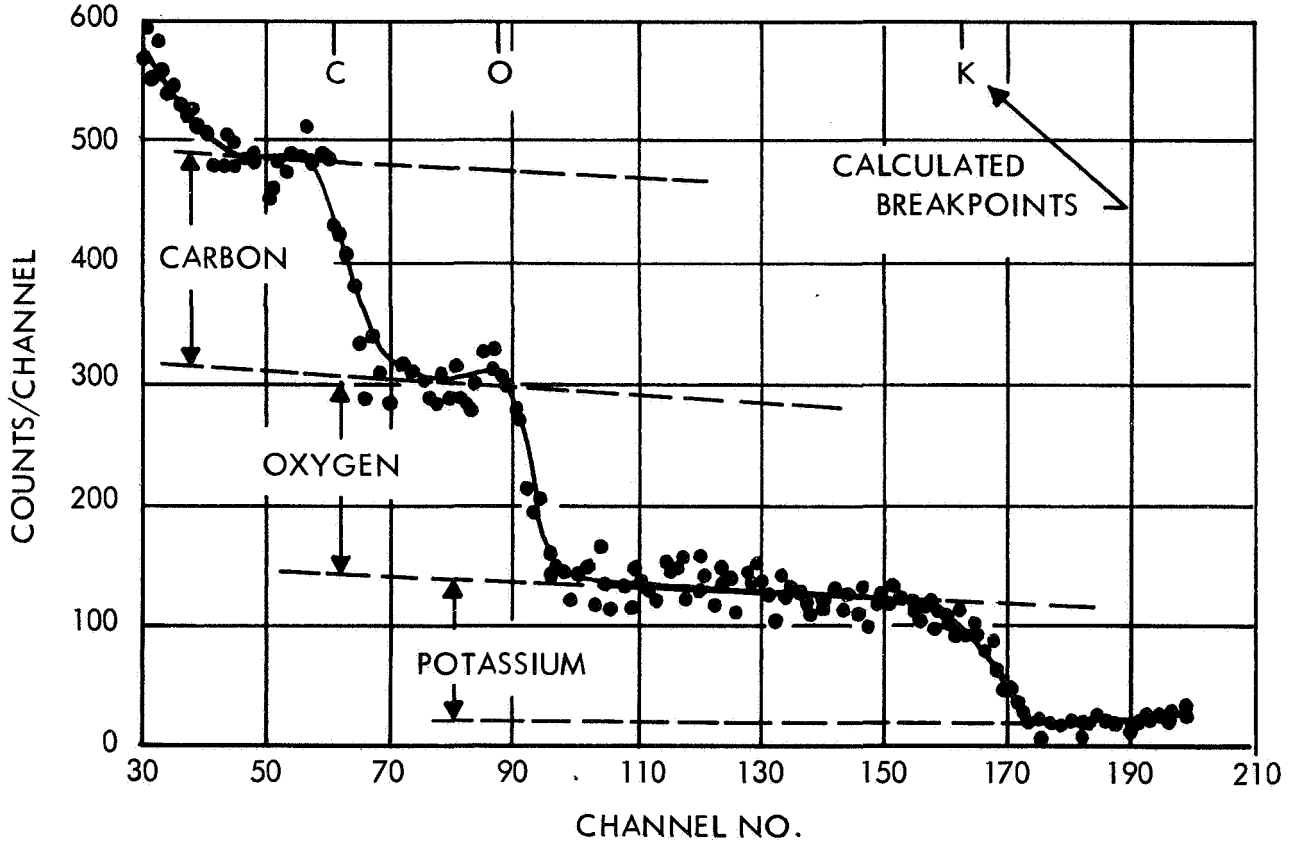


Figure A3. K_2CO_3 Alpha Spectrum (Reference 17)

Spectrometer Experiments -- The descent and entry phase is expected to provide a situation where on-board decision-making from spectrometer output is required. It might for instance be necessary to analyze the output of an ultra-violet spectrometer to determine what range of atomic weights the mass spectrometer should scan. This selection of a specific range for the mass spectrometer is needed due to the lack of time in communicating results from the lander to the orbiter prior to impact. This would be essential for a hard lander and would be desirable for a soft lander in case the landing was unsuccessful.

In general the output of such an instrument will be a signal - no signal answer for each frequency interval. This might be stored in digital form by recording the position along the frequency scale of each change in signal.

The processing task is to determine if the spectral lines of a known element are included in the spectral lines of the generated spectrum.

Picture Processing

Given the capability to examine the contents of a picture, a number of decision-making requirements can be fulfilled. Some of the situations requiring decisions are as follows:

1) In a mapping experiment there is a need to identify an area of special interest so that increased resolution, filter changes, or telescopic zooming might be tried. Such an identification must be made quickly before the object is out of the field of view.

2) A quick look at the intensity range of a picture might indicate that some changes in the picture taking procedure is required. It would also indicate when a picture contains little information, and is not worth further processor or transmission to earth.

3) This requirement for a decision is concerned with data compression. Regions of a picture (4 x 4 points, for instance) might be processed to determine the average intensity as well as the minimum and maximum intensity. If the region appears to have little information, simply send the average intensity and the identity of the block. If the region has much interest, send all the points.

4) In the case of a roving vehicle exploring the surface of the planet, the object identification capability can be used to identify objects for close examination and also to pass-up objects similar to some already examined.

The picture processing requirement is handled by procedures that vary from very simple to rather complex computations. For instance a simple procedure to identify potential objects of special interest is to identify areas of sharp contrast by simply comparing the intensity of neighboring points. The more sophisticated techniques would be to use the operator methods previously described. In this case, however, in addition to generating a characteristic vector for the input pattern, a comparison must be made with a number of stored vectors. This comparison is done by determining the distance between the generated characteristic vector and all the stored vectors, and then determining the minimum of these distances to identify the picture.

Other Decision Making Situations

Allocation of Resources -- While adaptive assignment and control of sensors have been described in the Data Handling section, a similar control function is required on other resources. One example is power. There is a need to match the solar cell array to the electrical load requirements. There is the need for adaptivity since solar cells may become disabled and load requirements will vary from one phase to the next and from one situation to another.

Task Scheduling -- The scheduling of experiments and amount of time spent on each can be expected to vary depending on the conditions encountered. This cannot be completely predetermined since the environment which the vehicle will operate is largely unknown.

Miscellaneous Processing Tasks

There exists a number of specific data processing tasks which do not conveniently fit into any of the previous categories.

Side Looking Radar Processing Requirements

Side looking radar is useful in planetary exploration to obtain high resolution. All weather pictures which cannot be obtained by using the visual spectrum. The main advantage of a side looking radar over conventional radar is that high resolution pictures can be obtained even though the antenna is small. The computational requirements, however, are severe.

In a conventional radar system high resolution is obtained by using large antennas. The side-looking radar system simulates a large antenna by processing the data from a small antenna as it is in successive positions due to the motion of the vehicle upon which it is mounted (Figure A4).

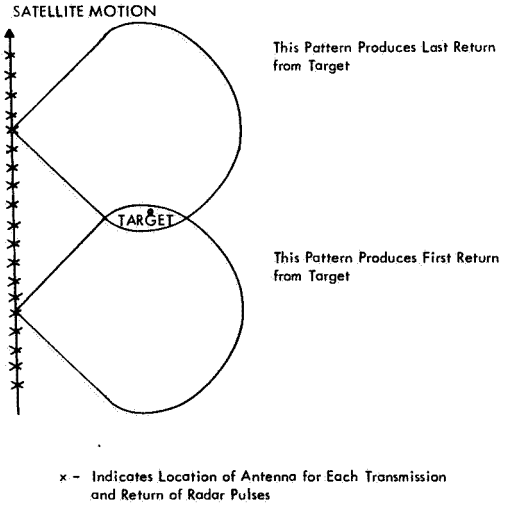


Figure A4. Side-Looking Radar Pattern

The signal returns from a target at successive positions of the antenna are stored and later processed to obtain the equivalent of the summing operation that is inherent to a large antenna.

For a satellite with a side-looking radar an antenna small in terms of wave length is mounted on the side of the satellite. The antenna beam is directed normal to the direction of motion of the vehicle. The radar scans out in range covering a swath of the planet of on the order of 50 miles on the side of the satellite. The returns from each increment of range (range bin) are stored. The size of the range bin is determined by the resolution desired from the radar. For a 50 mile swath and a 1/4 mile resolution there will be 200 range bins and sets of inputs per transmitted pulse. Signals are transmitted with a pulse repetition frequency that is a function of swath width and other parameters.

It is necessary to store the returns from the set of successive transmissions that illuminate a point on the ground (these correspond to the returns at different parts of a large physical antenna) and by computation perform the summation on the total set of returns. Each returned signal is utilized for determining the output for the set of points at that range covered by the beam as it moves. Therefore, one must store the returns for the time a beam covers each point rather than processing them as they are received and then forgetting them. Since the returned signal is in the form of a vector having amplitude and phase it is necessary to store two signal components for each return. Also, a compensation must be made for the differential phase shift that occurs for the different positions of the antenna with respect to the point being observed. The compensation is made because of the difference in path length to a given target from the different transmission points (Figure A5).

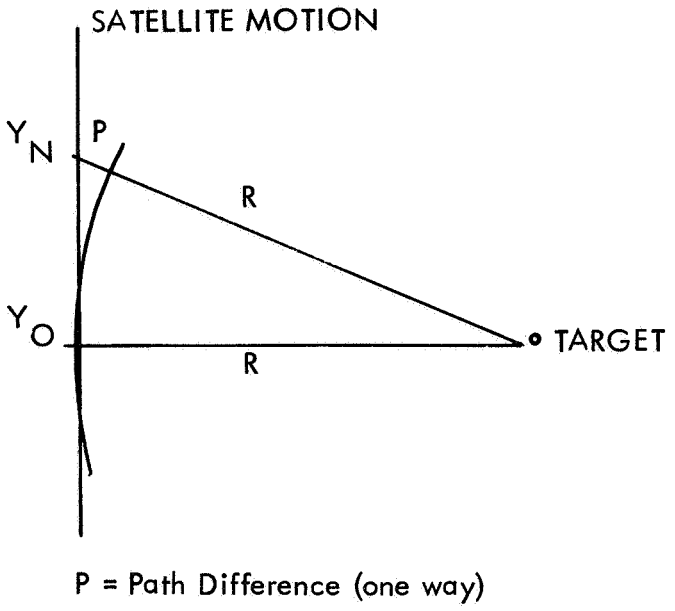


Figure A5. Geometry For Side-Looking Radar

A correction is made by shifting the phase $\frac{2P}{c}\omega_o$ radians, where P is the one way path difference, ω_o is the radar transmission frequency and c is the velocity of light. This requires storage of the phase shift for each point used in the processing.

In the processing operation the received signal is phase shifted properly prior to the signals being summed, then the summation takes place. The compensation for phase shift allows one to utilize all returns in which the target point was within the beam width of the radar rather than limiting the number used to those with only a small phase shift. This is the key to the high resolution of the SLR.

The following derivation is taken from Reference 17. The equations that must be solved for each output point are:

$$Z(R, Y_N) = \sum_{K=-M}^{M} \phi(R, N-K)\, S(R, N-K) \qquad (1)$$

$$R = 1, 2, \ldots T$$

$$N = 1, 2, \ldots$$

Where:

R is the number of the range bin for the point being determined,

T is the number of range bins,

Y_N is the along track coordinate of the point,

S (R, N-K) is the vector signal return in range bin R for the K-th transmission point prior to Y_N,

ϕ (R, N-K) is the differential phase shift vector for the position (R, Y_N) when the return taken at (R, Y_K) is used, 2M+1 is the number of returns being utilized in the computation.

For digital computation, the received signals would be converted from phase and amplitude to in-phase and quadrature components so that (1) would become

$$Z(R, Y_N) = \sum_{K=-M}^{M} \left[\alpha(R, N-K) + i\beta(R, N-K)\right] \times \left[U(R, N-K) + i\, V(R, N-K)\right]$$

$$= \sum_{K=-M}^{M} (\alpha U - \beta V) + i\,(\alpha V + \beta U) \qquad (2)$$

Using equation (2) the SLR requires the following operations:

a) Conversion of the received signals to in-phase and quadrature components and storage (handled by analog techniques).

b) Four multiplications and four additions for each of the 2M + 1 returns used in obtaining each output point.

The data storage requirement (a) is that of storage of the two components of the phase shift vector for (M + 1)T points and (2M + 1)T input value where T is the number of range values (200). In order to provide near real time processing the computing must be done for all T range values in the time required to move one resolution element. For a resolution element of 1/4 mile and a satellite velocity of 16, 400 ft/sec. (Venus orbital velocity), the allowable time is 1320/16, 400 = .0805 sec for the T points. Assuming sequential processing on the T points this would require

$$\frac{4 \times 200\ (2M + 1)}{0.0805} \approx 10^4\ (2M + 1)$$

multiplications and a like number of additions per second. For a typical value of M of 200 this is then on the order of 4×10^6 multiplications and additions per second.

The computations operate on low precision information. Both the returned signals and the phase shifts can be digitized with only a few information bits each.

NAVIGATION OF SPACECRAFT

That portion of the navigation problem considered here is concerned with computation of a reference trojectory, computation of transition matrices, and filtering of observations in order to determine the present state of the vehicle.

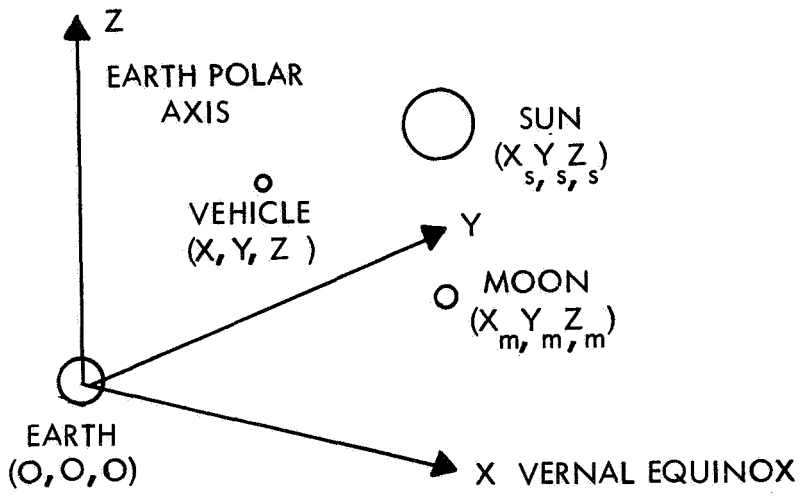
The reference trajectory generation methods considered are applicable to missions employing high-thrust corrective maneuvers. The class of low-thrust spiral trajectories consisting of several hundred revolutions, which are applicable in the departure and approach phases of a space mission, are not considered here. The data processing filters considered for navigational purposes are capable of yielding state (position, velocity) estimates from self-contained optical-inertial systems as well as ground supported navigation systems. The techniques for attitude determination and guidance corrections are not considered in this section.

Reference Trajectory and State Transition Matrix Computational Techniques

Reference Trajectory - Equations of Motion

The equations of motion which represent the space vehicle's motion naturally depend on the mission to be flown. Even for a specific flight, approximations are made in the equations representing the reference trajectory. The equations of motion are presented here on the basis of including the gravitational effects on the vehicle of the earth (including the second harmonic term of the earth's oblateness) and a spherical and homogenious moon and sun.

The coordinate system chosen is that of a nonrotating Cartesian geocentric frame. The Z axis lies along the earth's polar axis, positive to the north. The X and Y axes are in the equatorial plane with the positive X asis in the direction of the first point of Aries and the Y axis oriented so as to produce a right-handed orthogonal system. A diagram of this coordinate system is given in the accompanying sketch.



The equations of motion expressed in the coordinate system described are as follows:

$$\ddot{X} = -\frac{\mu_e X}{r^3}\left[1 + J\left(\frac{a}{r}\right)^2\left(1 - 5\frac{Z^2}{r^2}\right)\right] - \frac{\mu_m(X - X_m)}{\Delta_m^3} - \frac{\mu_m X_m}{r_m^3} - \frac{\mu_s(X - X_s)}{\Delta_s^3} - \frac{\mu_s Y_s}{r_s^3}$$

$$\ddot{Y} = -\frac{\mu_e Y}{r^3}\left[1 + J\left(\frac{a}{r}\right)^2\left(1 - 5\frac{Z^2}{r^2}\right)\right] - \frac{\mu_m(Y - Y_m)}{\Delta_m^3} - \frac{\mu_m Y_m}{r_m^3} - \frac{\mu_s(Y - Y_s)}{\Delta_s^3} - \frac{\mu_s Y_s}{r_s^3}$$

$$\ddot{Z} = -\frac{\mu_e Z}{r^3}\left[1 + J\left(\frac{a}{r}\right)^2\left(3 - 5\frac{Z^2}{r^2}\right)\right] - \frac{\mu_m(Z - Z_m)}{\Delta_m^3} - \frac{\mu_m Z_m}{r_m^3} - \frac{\mu_s(Z - Z_s)}{\Delta_s^3} - \frac{\mu_s Z_s}{r_s^3}$$

where

$$r = \sqrt{X^2 + Y^2 + Z^2}$$

$$r_m = \sqrt{X_m^2 + Y_m^2 + Z_m^2}$$

$$r_s = \sqrt{X_s^2 + Y_s^2 + Z_s^2}$$

$$\Delta_m = \sqrt{(X - X_m)^2 + (Y - Y_m)^2 + (Z - Z_m)^2}$$

$$\Delta_s = \sqrt{(X - X_s)^2 + (Y - Y_s)^2 + (Z - Z_s)^2}$$

$\mu_e = 3.986135 \times 10^{14}\ m^3/sec^2$

$\mu_m = 4.89820 \times 10^{12}\ m^3/sec^2$

$\mu_s = 1.3253 \times 10^{20}\ m^3/sec^2$

a = radius of earth at equator = 6.37826×10^6 m

$J = 1.6246 \times 10^{-3}$

The first, second, and fourth terms on the right side of the above three equations represent the gravitational attraction upon the vehicle of an oblate earth (second harmonic only), a spherical moon, and a spherical sun, respectively. The third and fifth terms represent the influence of the moon and sun upon the earth.

State Transition Matrix - Pertubation Equations

The transition matrix, Φ, can be computed by a number of methods. For largely three dimensional trajectories (e.g. translunar), Φ can be computed by solving the linearized pertubation equations of motion. These perturbation equations are obtained by linearizing the vehicle equations of motion.

The equations of vehicle motion are of the form:

$$\left.\begin{aligned} \ddot{X} &= f_1 (X, Y, Z) \\ \ddot{Y} &= f_2 (X, Y, Z) \\ \ddot{Z} &= f_3 (X, Y, Z) \end{aligned}\right\}$$

To linearize these equations, we expand each in a Taylor series about a reference position, X_R, Y_R, Z_R, for example,

$$\ddot{X} = f_1 (X_R, Y_R, Z_R) + \frac{\partial f1}{\partial X} (X - X_R) + \frac{\partial f1}{\partial Y} (Y - Y_R) + \frac{\partial f1}{\partial Z} (Z - Z_R) + \text{higher}$$

order terms and similarly for the $\ddot{Y}$ and $\ddot{Z}$ equations. Here it is understood that the partial derivatives are evaluated at the reference position. If the higher order terms are dropped (a reasonable approximation when the difference quantities $X - X_R$, etc., are small), the equations are linear in the difference quantities.

It is convenient to describe the state of this system of dynamical equations in terms of the difference quantities, remembering that X, Y, Z, X_R, Y_R, Z_R are all functions of time.

Thus, the state is a six vector:

$$\{\overline{x}\} = \begin{Bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{Bmatrix}$$

where

$$\left.\begin{aligned} x_1 &= X - X_R \\ x_2 &= Y - Y_R \\ x_3 &= Z - Z_R \\ x_4 &= \dot{X} - \dot{X}_R = \dot{x}_1 \\ x_5 &= \dot{Y} - \dot{Y}_R = \dot{x}_2 \\ x_6 &= \dot{Z} - \dot{Z}_R = \dot{x}_3 \end{aligned}\right\}$$

The equations of vehicle motion are then of the form

$$\left.\begin{aligned} \ddot{x}_1 &= \frac{\partial f1}{\partial X} x_1 + \frac{\partial f1}{\partial Y} x_2 + \frac{\partial f1}{\partial Z} x_3 \\ \ddot{x}_2 &= \frac{\partial f2}{\partial X} x_1 + \frac{\partial f2}{\partial Y} x_2 + \frac{\partial f2}{\partial Z} x_3 \\ \ddot{x}_3 &= \frac{\partial f3}{\partial X} x_1 + \frac{\partial f3}{\partial Y} x_2 + \frac{\partial f3}{\partial Z} x_3 \end{aligned}\right\}$$

These three second-order differential equations are now rewritten as the following six first-order differential equations:

$$\dot{x}_1 = x_4$$

$$\dot{x}_2 = x_5$$

$$\dot{x}_3 = x_6$$

$$\dot{x}_4 = \frac{\partial f1}{\partial X} x_1 + \frac{\partial f1}{\partial Y} x_2 + \frac{\partial f1}{\partial Z} x_3$$

$$\dot{x}_5 = \frac{\partial f2}{\partial X} x_1 + \frac{\partial f2}{\partial Y} x_2 + \frac{\partial f2}{\partial Z} x_3$$

$$\dot{x}_6 = \frac{\partial f3}{\partial X} x_1 + \frac{\partial f3}{\partial Y} x_2 + \frac{\partial f3}{\partial Z} x_3$$

or in matrix notation

$$\dot{\bar{x}}(t) = F(t)\,\bar{x}(t)$$

where $\bar{x}(t)$ is a six vector and $F(t)$ is a (time-varying) 6 x 6 matrix defined as follows:

$$F = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ \frac{\partial f1}{\partial X} & \frac{\partial f1}{\partial Y} & \frac{\partial f1}{\partial Z} & 0 & 0 & 0 \\ \frac{\partial f2}{\partial X} & \frac{\partial f2}{\partial Y} & \frac{\partial f2}{\partial Z} & 0 & 0 & 0 \\ \frac{\partial f3}{\partial X} & \frac{\partial f3}{\partial Y} & \frac{\partial f3}{\partial Z} & 0 & 0 & 0 \end{bmatrix}$$

If the partials in F are evaluated along the reference trajectory, they are merely functions of time once this reference has been selected. If the reference for which the partials are evaluated is the estimated trajectory, then the partials must be computed as the flight progresses since the estimate is not known a priori and is subject to change as each observation is made.

The above set of six first order perturbation equations are solved when any set of six linearly independent solutions is known. Let such a set of solutions be designated as separate columns of the matrix with elements x_{ij}. Since any linear combination of these columns also gives a solution, the columns of

$$\Phi(t_o, t) \triangleq \left[x_{ij}(t)\right] \left[x_{ij}(t_o)\right]^{-1}$$

must all be solutions. The transition matrix $\Phi(t_o, t)$ is equal to the identity matrix when $t = t_o$; this provides the necessary initial conditions to find its components by numerical integration. Since each of the columns of $\Phi(t_o, t)$ satisfy the above first order perturbation equations, Φ itself must satisfy

$$\dot{\Phi} = F\,\Phi$$

$$\Phi(t_o, t_o) = I$$

where F is a 6x6 matrix of the form

$$F = \begin{bmatrix} 0 & 1 \\ F^* & 0 \end{bmatrix}$$

and F* is the 3x3 gravitational gradient matrix which can include the effects of oblateness, solar-lunar, and gradient of the primary gravity field.

General Representation Methods for the Reference Trajectory and State Transition Matrix

The following trajectory and state transition matrix representation methods are felt to be feasible for on-board use:

i) numerical integration
ii) tabular interpolation - extrapolation
iii) analytic two-body expressions

A brief description of the requirements, advantages, and disadvantages of these methods follows.

The numerical integration techniques involve the integration of a large number of differential equations. For example, the reference trajectory must be represented by a set of six first-order differential equations. The six perturbation differential equations for the state transition matrix must be solved six times per given data point. The mechanization of a set such as this and the associated channeling logic which would be required impose a severe burden on the on-board computer. However, numerical integration has the advantages of being a very flexible and accurate method.

An alternative technique is the tabular technique. Here interpolation-extrapolation formulae are used with tabular entries for representation. The representation equations take the simple form of algebraic polynomials in time, and for this reason, the formulae are restricted to short prediction times and "flat" trajectories. The computation loads in this case also do not appear to be simple. Here large memory banks are required to initialize the interpolation-extrapolation formulae. For example, just to represent the transition matrix, a five-point extrapolation formula would require a set of five matrices of thirty-six element. This set would have to be frequently re-initialized with new sets of tabular entries. Compounded with the foregoing are the problems of generating the inverses and the reference trajectory itself. It is easy to see that large memory loads are implied in this method.

Other techniques for representation which have promise are the analytic methods. One technique is a closed-form first-order pertubation on the two-body integrals. In this technique, the effects of perturbations can be added to the two-body representation to make it more accurate. The advantage of this method is that the trajectory parameters and the transition matrices can be computed from closed form expressions at any particular time. Another advantage is that this method is ideal for fixed memory computers as it does not require prior data point storage. From the appearance of the expressions, it is not clear, however, that the computation loads are measurably less in this case over the numerical integration methods. The computation logic is more complex, but the computation itself is simpler. The numerical integration method suffers from the accumulation of round-off and truncation errors, while the analytic method suffers from truncation in the formula.

Table A1 is a summary of the computational load characteristics for the above general representation techniques for reference trajectories and their associated state transition matrices.

State Estimation Computations

Navigation is assumed to be performed by use of an augmented inertial system (e.g., a pure inertial system augmented by two star trackers and horizon scanner). Observations can be assumed to be made every 5 to 10 minutes during observational periods, with allowable periods (say 30 to 60 minutes) during the mid-course phase in which no observations are taken. A typical augmented inertial system using platform mounted telescope and a gimbaled inertial platform with gyro and accelerometer package yields 2 simultaneous angular measurements per observation.

Table A1. Computation Loads Summary for Trajectory and Transition Matrix Representation Theories

Representation Theory	Computation Load	Memory Load	Comments	
			Advantages	Disadvantages
1. Numerical Integration of Equations of Motion	1. Large 2. Moderately complex logic	1. Essentially none, if initialized from ground 2. Moderate if initialized from decision points 3. Store trajectory parameters at decision points	1. Versatile 2. Can handle abort easily 3. Can be made very accurate	1. Complex integration for each data point 2. Complex channeling logic 3. Accumulation of roundoff and truncation
2. Tabular Interpolation Extrapolation	1. Moderate - mostly channeling logic 2. Simple logic	1. Very large 2. Tabular store trajectory parameters and transition matrices 3. Abort conditions require additional, large banks	1. Simple mechanization 2. Can be very accurate over prediction time	1. Approximate 2. Good for short prediction times or "flat" trajectories initialization
3. Analytic Two-Body Integrals	1. Large 2. Complex logic	1. Moderate if initialized from decision points 2. Store trajectory parameters at decision points	1. Good for fixed memory computer 2. Versatile 3. Result computed directly for each data point	1. Complex logic 2. Requires "Patching" computation 3. Approximate 4. Logic must be changed for different logics

There are two basic approaches to navigation via optimal linear filtering. The first method consists of processing each observation as it is received and is referred to as recursive. Kalman's theory is in this category. A second non-recursive method used the batch data approach. In the latter technique observational data is stored and processed all at one time. Basic least squares is an example of this approach.

Either of the above two approaches can be used in the primary mission phases commonly referred to as departure, orbital, mid-course, and approach. However, only the Kalman filter will be considered here.

Mathematical Model of Navigation Problem

In essence, the navigation problem is a problem of optimal estimation of a linear sampled data system. Reduced to its essentials, this problem can be shown to be characterized by the following two equations:

$$y_n = M_n x_n \qquad \text{(Trajectory Determination Equation)}$$

$$x_{n+1} = \Phi_{n+1,\,n}\, x_n \qquad \text{(Trajectory Propagation Equation)}$$

where

y_n = observation residuals vector for the n^{th} (current) navigation cycle

x_n = position-velocity deviation state vector at the beginning of n^{th} cycle (state at t_n)

M_n = matrix relating current position-velocity deviation to observation residuals

x_{n+1} = deviation state at n = 1 (prediction)

$\Phi_{n+1,\,n}$ = state transition matrix between two points, n+1 and n

Hereafter, the words deviation state and position-velocity will be used interchangeably. In general, x_n is a 6 x 1 column vector, y_n is P x 1 column vector, M_n is P x 6 matrix and $\Phi_{n+1,n}$ is 6 x 6 matrix.

State Determination

In the following, techniques by which the observational data (e.g., angles between celestial bodies) is processed to yield estimates of the trajectory deviation parameters (position and velocity) from their reference values are considered.

The trajectory estimation problem is to determine the six coordinates of position and velocity deviation x_n from the trajectory determination equation. If the data residual vector y_n has completely independent coordinate measurements and is 6 x 1, then the parameters can be determined simply by inverting the trajectory determination equation. This is referred to as the exactly-determined case.

The general problem is that y_n may be less than 6 x 1 (under-determined), or it may be greater than 6 x 1 (over-determined). In the former case parameter estimation techniques (Kalman optimal filter theory) are used to determine the deviation state vector. In the latter case, the method of least squared, the maximum likelihood principle, or the Bayes Estimator are often used to determine the trajectory parameters.

Given a set of observation residuals, y_n, having components equaling the number of state elements, under-determined or over-determined, the problem is to compute an estimate of the current deviation state x_n which measures the present position-velocity deviation from the reference values. Knowing the reference trajectory parameter values and the

deviation state, an estimate of the desired actual state of the spacecraft is obtained. This problem is called trajectory estimation.

Because of measurement and resolution errors in the observation residuals, and also because it may be desirable to base the estimate on an incomplete set of observations, statistical methods are required to optimally extract the deviation state x_n from the set of observation residuals y_n. This problem of optimal determination of the deviation from noisy observation is defined as the optimal filtering problem. The optimal estimate is defined by $\hat{x}_n$ (Optimal estimate is designated by the hat ($\wedge$) symbol in the following discussions).

Trajectory estimation based upon optimal filtering techniques can handle the under-determined as well as the over-determined trajectory determination equations. Its particular value lies in being able to handle the under-determined problem.

As it turns out in the theory, the under-determined trajectory estimation problem is the more general problem. The over-determined case, usually handled as a least squares or maximum likelihood problem, is a special case of Kalman's optimal filter.

Tables A2 and A3 give a set of basic computation routines for the trajectory estimation problem which is set for solution at the n^{th} decision point. These equations are based on Kalman's optimal filtering theory.

Table A2. Summary of Navigation Equations

1. Compute (for simulation purposes only) actual observation residual $\tilde{y}_n$ from actual state deviation x_n^A and Jacobian Matrix M_n^A evaluated with x_n^A

$$\tilde{y}_n = M_n^A \; x_n^A$$

2. Compute transition matrix

$$\Phi_{n,\, n-1}$$

3. Compute predicted deviation at n

$$\hat{x}'_n = \Phi_{n,\, n-1} \; \hat{x}_{n-1}$$

4. Compute covariance matrices

$C(\Delta \hat{x}'_n)$ and $C(\Delta \tilde{y}_n)$ (See Table A3)

5. Compute optimal data process filter ω_n

$$\omega_n = C(\Delta x'^{\wedge}_n)\, M_n^T \left[M_n \, C(\Delta x'^{\wedge}_n)\, M_n^T + C(\Delta \tilde{y}_n) \right]^{-1} \text{ (Kalman)}$$

6. Compute optimal estimate to current deviation state

$$\hat{x}_n = (I - \omega_n M_n) \; \hat{x}'_n + \omega_n \tilde{y}_n$$

7. Predict future deviation state

$$\hat{x}'_{n+1} = \Phi_{n+1,\, n} \; \hat{x}_n$$

Covariance Computation

It next remains to map computational methods for calculating the covariance matrices for the data processing filter. Table A3 is a summary of the computational routines for computing the covariance matrices. The covariance of measurement and resolution errors of the observation residuals, $C(\Delta \tilde{y}_n)$, is assumed to be known for the n^{th} cycle. The dependence of the elements of this matrix upon location in the trajectory, and the sensitivity of errors to

The difference yields

$$\Delta\hat{x}_n = \hat{x}_n - \hat{x}_n^A = \omega_n(\tilde{y}_n - \hat{y}'_n) + \Delta\hat{x}'_n$$

which can be simplified

$$\Delta\hat{x}_n = (I - \omega_n M_n)\,\Delta\hat{x}'_n + \omega_n\,\Delta\tilde{y}_n \quad \text{(Error in Estimation of Actual Deviation)}$$

where $\Delta\tilde{y}$ represents the measurement and resolution errors.

Table A3. Summary of Computational Routines for Covariances

1. Compute covariance of optimal estimate error at n-1

$$C(\Delta\hat{x}_{n-1}) = C(\Delta\hat{x}'_{n-1}) - \omega_{n-1}\left[M_{n-1}C(\Delta\hat{x}'_{n-1} + C(\Delta\tilde{y}_{n-1})\right]\omega_{n-1}^T$$

where

$C(\Delta\hat{x}'_{n-1})$ = covariance of predicted state error at (n-1), computed.

$C(\Delta\tilde{y}_{n-1})$ = covariance of measurement, and resolution errors of observation residual.

M_{n-1} = computed from geometry and reference or estimated trajectory data.

ω_{n-1} = optimal weighting matrix, computed.

$C(\Delta\hat{x}'_o)$ = covariance matrix of injection errors.

ω = 0 = (no data process at this time.)

2. Compute covariance of error in optimal prediction

$$C(\Delta\hat{x}'_n) = \Phi_{n,\,n-1}\,C(\Delta\hat{x}_{n-1})\,\Phi^T_{n,\,n-1}$$

where $\Phi_{n,\,n-1}$ is evaluated with the reference trajectory parameters.

State Prediction

Another aspect of navigation problem is prediction. Having extracted an optimal estimate of the present state $\hat{x}_{n,}$ it is necessary to optimally predict

particular directions of the observed celestial bodies, can be taken into account by way of this matrix. The error sources which might be considered are platform or vehicle altitude drift rates, pointing resolution errors, pointing readout errors, and unaccounted trajectory perturbation errors. Measures of $C(\Delta\tilde{y}_n)$ for each n can be pre-computed or calculated on board using parameter values of the reference trajectory.

With regard to the covariance of state estimation errors, $C(\Delta\tilde{x})$, we have:

$$\Delta\hat{x}'_n = \hat{x}'_n - x^A_n = \Phi_{n,n-1}(\hat{x}_{n-1} - x^A_{n-1})$$
$$= \Phi_{n,n-1}\,\Delta\hat{x}_{n-1}$$

This equation states that the error in the predicted estimate of the state $\hat{x}'_n$ from the actual state x^A depends on errors of the optimal estimate at n-1 of the actual state.

From the above expression the covariance of $\Delta\hat{x}'_n$ can be computed from $\Delta\hat{x}_{n-1}$

$$C(\Delta\hat{x}'_n) = \Phi_{n,n-1}\,C(\Delta\hat{x}_{n-1})\,\Phi_{n,n-1} \quad \text{(Covariance of Error in Optimal Prediction)}$$

This equation gives the covariance of errors of the present predicted state in terms of the covariances of past errors. To initialize this computation with respect to the first navigation cycle, it should be noted that $C(\Delta\hat{x}_o) = C(\Delta\hat{x}'_o)$ = covariance of injection errors which is assumed to be known and $\omega_o = 0$, implying that no observation data is processed at injection time.

To complete the calculation, it is necessary to compute the covariance $C(\Delta\hat{x}_{n-1})$. This term can be developed for the n^{th} cycle in the following way. Recalling from Table A2,

$$\hat{x}_n = (I - \omega_n M_n)\,\hat{x}'_n + \omega_n(\tilde{y}_n)$$
$$x^A_n = \Phi_{n,n-1}\,x^A_{n-1}$$

the future state of the deviation for making guidance decisions. This constitutes the optimal prediction problem. The optimal state prediction at n+1 is

$$\hat{x}'_{n+1} = \Phi_{n+1,n}\,\hat{x}_n$$

where the prime is used to denote that prediction is made without observational data between n+1, and n, and only upon the most recent optimal estimate. This equation can be used to arrive at a decision of whether or not to institute further data gathering or velocity correction at n.

State prediction is shown as the last computation in Table A2.

NAVIGATION AND CONTROL OF LANDER

The terminal guidance sub-system for the lander is assumed to consist in part of a strapped-down navigation system employing a set of laser gyros as inertial rate sensors. It is expected that the use of laser gyros will provide high precision input data with high reliability and at a relatively low cost. Some of the outputs of portions of the navigation computer are also used as inputs to the stabilization and control sub-system.

Description of Requirement

A strapped-down or gimballess inertial navigation and control system of this type employs the laser gyros and a set of direction cosine computations to establish the space-fixed coordinate system in which acceleration is integrated twice to yield position. The laser gyros and the direction cosine computation therefore perform, in an analytic sense, the same function that the gimbals perform in a gimballed navigation system and the coordinate transformation performs the same function that the platform performs in a gimballed system. The remainder of the navigation computation is essentially the same for either type of system. The stabilization and control computations employ the same set of direction cosines as used in the navigation computer.

The major computational tasks are separated and identified in block diagram form in Figure A6. The outputs of the laser gyros are sent first to the block marked gyro signal corrections which take on the following form.

GYRO SIGNAL CORRECTIONS

$$\begin{pmatrix} \Delta\theta_1 \\ \Delta\theta_2 \\ \Delta\theta_3 \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{pmatrix} \begin{pmatrix} \Delta\theta'_1 \\ \Delta\theta'_2 \\ \Delta\theta'_3 \end{pmatrix}$$

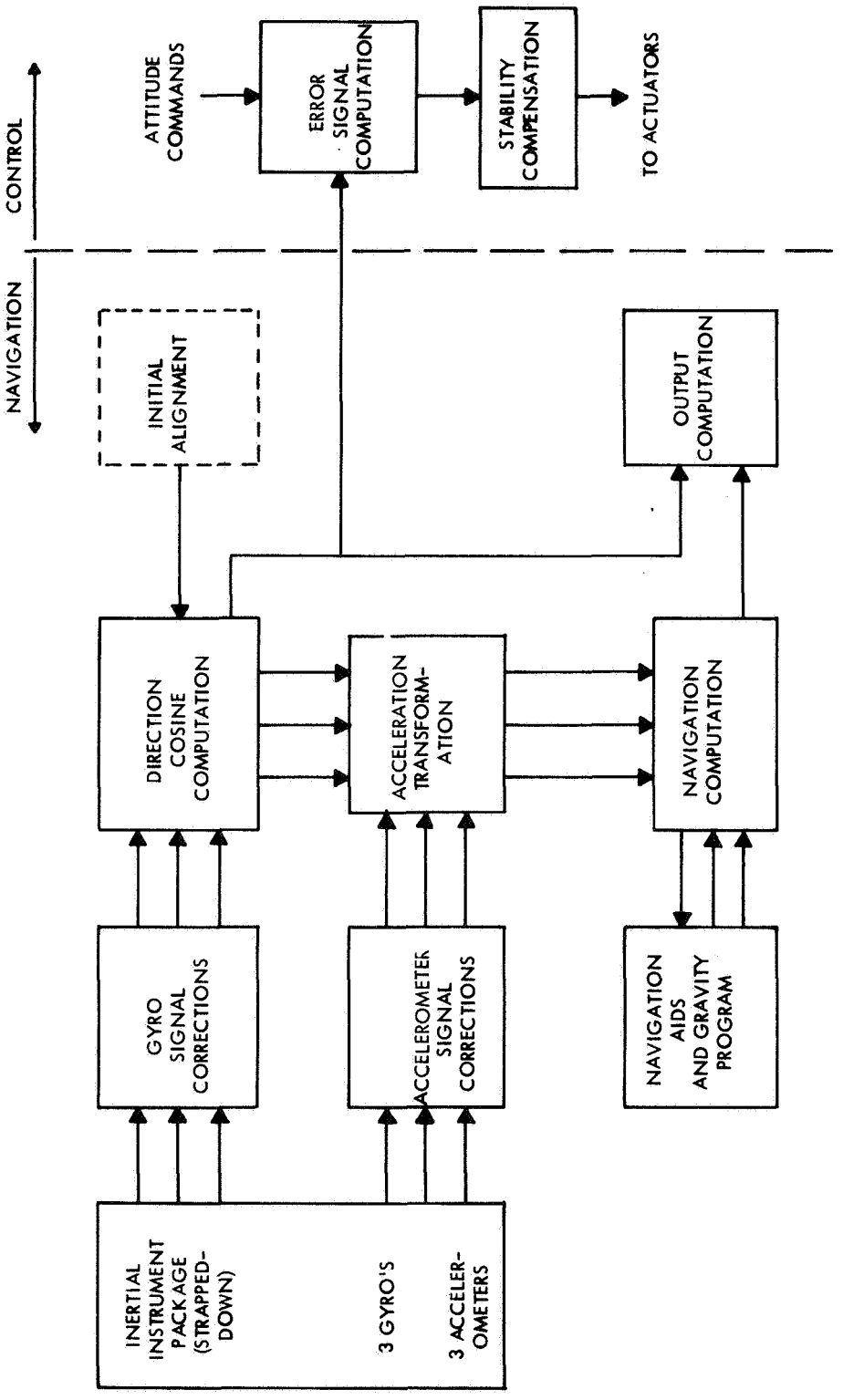


Figure A6. Navigation And Control of Lander

This computation corrects the gyro outputs for non-orthogonality and for variation in the scale factor between individual gyros. It also scales each increment such that they each have a value of 2^k radian, where k is an integer. In a particular situation it may be possible to omit some or all of these computations. Regardless of the number of corrections that are necessary all of them can be made simultaneously in the computer. These computations are in the form of a matrix multiplication.

Accelerometer corrections for variation in the scale factor for non-orthogonality, and for size effect are handled by the following equations:

ACCELEROMETER SIGNAL CORRECTIONS

$$\begin{pmatrix} \Delta V_1' \\ \Delta V_2' \\ \Delta V_3' \end{pmatrix} = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix} \begin{pmatrix} \Delta V_1'' \\ \Delta V_2'' \\ \Delta V_3'' \end{pmatrix}$$

$$\Delta V_1 = \Delta V_1' - S_1 K [(\Delta\theta_2)^2 + (\Delta\theta_3)^2]$$
$$\Delta V_2 = \Delta V_2' - S_2 K [(\Delta\theta_3)^2 + (\Delta\theta_1)^2]$$
$$\Delta V_3 = \Delta V_3' - S_3 K [(\Delta\theta_1)^2 + (\Delta\theta_2)^2]$$

The first two of these corrections can be mechanized as a matrix multiplication in a manner similar to that used for the gyro corrections. The size effect computation, however, is quite different and requires the use of gyro signals as well as those from the accelerometers. The size effect is a result of the centripetal accelerations sensed by the accelerometers due to the fact that they can not all be located at the same point in space.

A total of 9 direction cosines must be obtained. The required computations can be broken down into three identical sets of equations.

DIRECTION COSINE COMPUTATION

$$\ell_1 (n+1) = \ell_1 (n) + \Delta\theta_3 \ell_2 (n) - \Delta\theta_2 \ell_3 (n)$$
$$\ell_2 (n+1) = \ell_2 (n) + \Delta\theta_1 \ell_3 (n) - \Delta\theta_3 \ell_1 (n)$$
$$\ell_3 (n+1) = \ell_3 (n) + \Delta\theta_2 \ell_1 (n) - \Delta\theta_1 \ell_2 (n)$$

(Plus identical equations for $m_i (n+1)$ and $p_i (n+1)$

Each set consists of three difference equations, two of which must be solved each time an input increment occurs. The equations are solved in the given order if the input increment is positive and in the reverse order if the increment is negative. This reverible algorithm has been shown to minimize error build up in the generation of the direction cosine.

The acceleration transformation computation transforms the components of specific force sensed by the accelerometers from the vehicle coordinate system to the space fixed navigational system.

ACCELERATION TRANSFORMATION

$$\begin{pmatrix} \Delta V_1^* \\ \Delta V_2^* \\ \Delta V_3^* \end{pmatrix} = \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \\ p_1 & p_2 & p_3 \end{pmatrix} \begin{pmatrix} \Delta V_1 \\ \Delta V_2 \\ \Delta V_3 \end{pmatrix}$$

This computation is also in the form of a matrix multiplication.

The remaining navigation computation blocks are essentially the same whether the system is gimballed or gimballess. The navigation computation consists of a double integration of acceleration to obtain velocity and position.

NAVIGATION COMPUTATION

$$V_i\,(n+1) = V_i\,(n) - h\,\lambda^2\,X_i\,(n) + \Delta V_i^*(n)$$

$$X_i\,(n+1) = X_i\,(n) + h\,V_i\,(n+1)$$

where: $h = \Delta t$; $i = 1, 2, 3$

The gravity vector G, which is required in this computation, is obtained by means of a mathematical model of the local gravitational field.

The outputs of the navigation computation of the velocity and position components are measured along the axis of the space fixed coordinate system. These are not in general the output quantities that are desired. Instead similar quantities referenced to the local geographical coordinates are usually desired. The purpose of the output computer is to transform these quantities from space fixed coordinates to the local geographical coordinates. This computation can be designed to produce other output quantities of interest such as those related to the attitude of the vehicle.

The error signal computation accepts the direction cosine and the attitude commands and computes error signals which specify the difference between the actual and the desired attitude.

ERROR SIGNAL COMPUTATION

$$e(\phi) = p_i\,C_{21} + p_2\,C_{22} + p_3\,C_{23}$$

$$e(\theta) = p_i\,C_{11} + p_2\,C_{12} + p_3\,C_{13}$$

$$e(\psi) = m_i\,C_{11} + m_2\,C_{12} + m_3\,C_{13}$$

The error signals are then transmitted to the stability computation block which is an eighth order difference equation. The output of this block is in the form of commands to the control devices.

COMMAND AND COMMUNICATIONS PROCESSING

Command Processing

Command processing proposed for the Voyager system is described in Reference 9. This is felt to be typical of command processing on unmanned space missions and is summarized here.

Each command received by the space craft may be for immediate action or it may be stored, as indicated by a portion of the command word. If a command is stored, it includes a time tag which indicates when the command is to be carried out. A Present Time Register is provided on the space craft. This provides a continuous indication of elapsed time for the comparison of command time tags and for their execution. The retrieval of command from the memory is achieved by searching for the lowest of all stored time tags that are greater than the present time. When this command is located it it brought out for continuous comparison with the Present Time Clock. Thus during this waiting period, minimum power is expended, since the memory is not behind accessed.

A summary of the functions involved in command processing

1) Receive and verify commands
2) Determine whether command to be executed or stored
3) Execute commands
4) Retrieve stored commands for execution at appropriate time
5) For repetative commands, compute a new time tag after the command has been executed.

Communications

There are a variety of coding and decoding requirements on the telemetered data. While it is felt that these requirements might include considerable computation, difficulty in obtaining information and lack of time did not allow expansion of this funtion.

REFERENCES

1. Conceptual Design Studies of an Advanced Mariner Spacecraft, Volume II, Systems Analysis, Prepared by Research and Advanced Development Division Avco Corporation, 28 October 1964.

2. Study of Mars and Venus Orbiter Missions Launched by the 3-Stage Saturn C-1B Vehicle, EDP-139, Volume III, JPL, 31 December 1963.

3. Voyager Design Study, Volume II, Mission and System Analysis, Document No. 63SD801, 15 October 1963.

4. Voyager Design Study, Volume II, Scientific Mission Analysis, Avco RAD-TR-63-34, 15 October 1963.

5. Voyager Design Study, Volume IV, System Design, Document No. 63SD801, 15 October 1963.

6. Voyager Spacecraft System Study, (Phase II Saturn V Launch Vehicle), Final Report, Volume I, Summary, Document No. 64SD4376, 9 December 1964.

7. Voyager Spacecraft System Study, (Phase II Saturn V Launch Vehicle), Final Report, Volumes IIa and IIb, Document No. 64SD4376, 9 December 1964.

8. Voyager Design Study, Volume III, Part 1, Subsystem Design, Document No. 63SD801, 15 October 1963.

9. Voyager Design Study, Volume III, Part 2, Subsystem Design, Document No. 63SD801, 15 October 1963.

10. Isidore Eisenberger, Edward C. Posner, "Systematic Statistic Used for Data Compression in Space Telemetry", Technical Report No. 32-510, JPL, 1 October 1963.

11. C. M. Kortman, "Data Compression and Adaptive Telemetry", Presented at 1965 WESCON, August, 1965.

12. Space Programs Summary No. 37-20, Volume VI, Space Exploration Programs and Space Sciences, JPL, 30 April 1963.

13. D. R. Weber, "A Synopsis on Data Compression", Proceedings of the 1965 National Telemetering Conference, April, 1965.

14. Dr. Richard Simpson, "Buffer Control in Data Compression Systems for Non-Stationary Data, Proceedings of the 1964 National Telemetering Conference, June, 1964.

15. Lawrence W. Gardenhire, "Redundancy Reduction-The Key to Adaptive Telemetry", Proceedings of the 1964 National Telemetering Conference, June, 1964.

16. A. V. Balakrishman, "An Adaptive Non-linear Data Predictor", Proceedings of the 1962 National Telemetering Conference Volume II.

17. "Computer Associative Memory Study", Final Report, Contract AF 04(695)318, Prepared by TRW for Space Systems Division, 15 July 1964.

APPENDIX B

SURVEY OF INFORMATION STORAGE DEVICES SUITABLE FOR ASSOCIATIVE MEMORIES

CONTENTS**

*Indicates devices which have been seriously considered for associative memory application in the past, or exhibit substantial promise.

**The number in parenthesis after each item is a "confidence level" for the entire classification.

*Indicates devices which have been seriously considered for associative memory application in the past, or exhibit substantial promise.

**The number in parenthesis after each item is a "confidence level" for the entire classification.

APPENDIX B

SURVEY OF INFORMATION STORAGE DEVICES SUITABLE FOR ASSOCIATIVE MEMORIES

This appendix concerns information storage devices, not associative memories per se. Relatively little discussion of associative memory algorithms is included, and much of the material would be equally applicable to program/data memories for general-purpose computers. However, there is considerably more emphasis on non-destructive readout (NDRO) devices than there would be in a survey of devices suitable for use in program/data memories.

Since the objective is to explore components of use during approximately the next 20 years, in accordance with the intent of the contract the appendix is rather non-committal about economics - for these can change rapidly as a result of fabrication process breakthroughs such as "planar" semiconductor techniques. Hence, no dogma will be invoked which would enable certain seemingly inappropriate memory devices to be excluded, since new technology may result in an abrupt revival. As an example, Lear-Siegler Corporation, Grand Rapids, Michigan, has recently announced a display (and inherently memory) device best described as a two-dimensional, batch-fabricated array of tiny gas discharge tubes; at one jump this announcement raises the possibility of a "neon light associative memory," which would seem rather grotesque if taken out of context.

The "Table of Contents" comprises an outline of the memory classification used in this appendix. No attempt will be made to defend this outline dogmatically justifiable, logically satisfying, complete, or properly balanced; it simply imposes some degree of order on the large and chaotic collection of devices which may be considered as candidates for associative memory operation by working computer designers in the future. To some extent, this imposed order tends to evaporate under close scrutiny, as there are exceptions to almost any scheme of ordering; however, a computer designer who knows memory devices well enough to think readily of these exceptions probably does not need the outline to begin with. This classification scheme is believed to be a "useful" one, and if so that is enough excuse for it.

FOREWORD:

An attempt will be made here to indicate some very subjective and tentative ratings, for the devices discussed in this memo, with respect to their usefulness as information storage elements in associative memories and associative computers. It should be clearly understood that such an attempt is highly speculative since the time frame of interest is 1975-1985. The measure to be used is defined below and is called "confidence level":

A confidence level of 2 means the device is used in commercial systems of relatively recently design which are operating satisfactorily thus far, but with which there is not a large body of actual field experience.

A confidence level of 3 means that the device has been demonstrated in a prototype complete system which is not yet in production, and that the prototype appears to operate satisfactorily.

A confidence level of 4 means that the device has been satisfactorily operated under laboratory conditions in a prototype subsystem which would be a portion of a complete system, but that a complete system has not yet been developed.

A confidence level of 5 means that small-scale "breadboard" models of the device have been operated under laboratory conditions.

A confidence of 6 means that some physical effect, on which the operation of the device would presumedly be based, has been demonstrated under laboratory conditions and is well understood; but that not even a small-scale prototype of the device itself has yet been demonstrated.

A confidence level of 7 means that a physical effect on which the device would presumably be based has been predicted theoretically, but that the effect itself has not yet been demonstrated, even under laboratory conditions.

A confidence of 8 is assigned where a device has not yet reached any of the above stages.

Obviously there are "shades of grey" between any two confidence levels, since these definitions are in turn based on some undefined terms such as "prototype," "satisfactorily," and "under laboratory conditions." A confidence level is assigned to the class of devices covered in each subsection, in the "Table of Contents."

1.0 ACTIVE ELEMENTS

An active element for the purposes of this document is "any element capable of accepting an input signal and converting it into a power amplified output signal of a similar type."

At present there are at least three major categories of active elements which may be considered as plausible associative memory logic-storage elements: integrated circuits, cryotrons, and discrete tunnel diodes. Of these, integrated circuits are probably of the most current interest.

1.1 Integrated Circuits

An integrated circuit may be defined as any batch-fabricated combination of active and passive elements composed of semiconductor materials, on a single substrate. From the point of view of associative memories it is useful to subdivide integrated circuits further into five categories: monolithic field-effect devices, including insulated-gate ("MOS" or "MOST") devices; thin film field-effect devices, which are still in the experimental laboratory stage and are not commercially available; silicon-controlled switches; and tunnel diodes.

1.1.1 Monolithic Bipolar Devices

Monolithic bipolar device integrated circuits are now commercially available for many different requirements, and are currently being supplied in the U.S. at an annual sales rate of roughly 50 million dollars. It is possible to fabricate these circuits solely by means of the "planar" batch-fabrication technique originally developed for the manufacture of silicon transistors; planar fabrication techniques have now been extended to produce passive elements, and also other types of active elements. Alternatively, planar techniques may also be used in conjunction with thin film deposition techniques; by means of the latter, metal-film resistors and capacitors may be deposited on top of the "passivating" oxide layer on the surface of a silicon chip containing planar-produced active elements below the surface.

There are many different types of monolithic bipolar devices integrated circuits. A good guide to them is the book Introduction to Integrated Semiconductor Circuits, A.J. Khambata, John Wiley and Sons, Inc., New York, 1963. Page 128 of this book gives a complete summary, with an attempt at rating, of the principal types of existing monolithic bipolar integrated circuits. The two which would appear best from a conservative design view-point are the ones which Khambata refers to as LCDTL and TTL, i.e., "Load Compensated Diode Transistor Logic" and "Transistor Transistor Logic"; the best form of TTL is apparently the HLTTL ("High Level TTL") now being advertised.

In principle, one could design a logic-storage element for an associative memory using discrete components such as transistors, diodes, resistors, and capacitors, and then more or less directly translate this design component-by-component into a monolithic integrated circuit. In practice, there are a few pitfalls which arise because of the manner in which these discrete elements are obtained by integrated circuit. For instance, in order to make an integrated circuit diode,a typical procedure is to make a complete integrated circuit transistor and then not connect its emitter; the base-collector junction is then effectively the diode. A resistor would be fabricated as a single region of semiconducting material with appropriate doping, having connection terminals at each end. To obtain a capacitor, the capacitance of a P-N junction is deliberately exploited. The range of component values obtainable by these techniques is much more restricted than that available with discrete components, the tolerances are much looser, there are side effects not present in discrete components, and the interconnection problems are different. Thus a final integrated design for a given circuit is likely to be substantially, although not totally, different from a discrete design for the same circuit.

1.1.2 Monolithic Field Effect Devices

If a "channel" of relatively conductive semiconductor material is fabricated, within a chip of "intrinsic" semiconductor material having a much lower conductivity, and leads are attached at each end of the channel, at that point the device is a resistor. If now a third "gate" connection is fabricated such that voltage can be applied between it and the channel, but there is no conductive path from the gate to the channel, the device becomes an "insulated gate" field-effect transistor, also called a MOS or MOST (Metal-Oxide-Semiconductor Transistor); the usual geometry is that the gate contact is on top of the "passivating" silicon oxide layer. It is also possible to make field-effect devices using strictly monolithic diffusion techniques, in which the gate is isolated from the channel by a backbiased P-N junction: this type of field-effect was introduced earlier than the MOST type. However, MOST devices can apparently be made physically smaller than virtually any other type of active elements by a factor estimated as to be as high as 10 by some sources, and this consideration may be decisive in the long run. For instance one can now buy, from either of two vendors, 20-bit MOST shift registers in one semiconductor device package. Experimental laboratory models of similar shift registers have been made in lengths up to 100 bits. MOS "table-lookup circuits" (that is, non-alterable memory devices) apparently are also practical; at the present state of the art, sizes of up to perhaps 256 logic elements can be fitted into one device package.

A "field-effect transistor" is actually not a "transistor" at all in the familar sense of the word. Rather, it is a fundamentally separate type of active device, whose internal operating principles are altogether different from that of conventional or "bipolar" transistors, and whose circuit characteristics resemble those of a vacuum tube much more than those of a bipolar transistor. Although field-effect devices will probably not be able to operate at as high frequencies as bipolar transistors, their frequency performance is being rapidly upgraded. Existing devices have gain-bandwidth

products of up to 300 megacycles, which is almost an order of magnitude better than that of the field-effect devices available two years ago. Although this parameter is not an explicit measure of the speed of the device for switching application, switching speeds have shown a similar improvement.

Two references on MOST memories are: "MOS Integrated Circuits Save Time and Money," D. E. Farina and D. Trotter, Electronics, 4 October 1965, pages 84-95; and "Integrated MOS Transistor Random-Access Memory", J. D. Schmidt, Solid State Design, January 1965, Pages 21-25.

1.1.3 Thin-Film Field Effect Devices

Two major types of batch fabrication technologies, for both active and passive integrated circuit devices, have been in vigorous competition for the last several years: monolithic techniques, and thin-film techniques. It is substantially correct to state that monolithic techniques are the most natural for active circuit component types, and thin-film techniques are the most natural for passive component types. Obviously, one always needs both types of components in a circuit of any complexity, and therefore either the two technologies must be mixed, or else one of them must be applied to components for which the other is really more appropriate.

Historically, thin-film fabrication techniques arose first. They are used extensively to made compact, environmentally tough circuits for aerq-space applications; resistors, capacitors and interconnectors (component-to-component conductive leads) can be made to very good tolerances in this way. The problem has been that, until recently, there has not been any really good way to make active elements using strictly thin-film techniques. Hence, in order to provide active elements, it has been necessary to in some way attach individually produced monolithic transistor chips to a previously prepared thin-film circuit containing the required resistors, capacitors, and interconnectors.

Recently, however, it has become possible to make thin-film diodes and field-effect devices, using different semiconductor materials than are used in monolithic devices. One promising material is cadmium sulfide, which has such a high-energy band-gap that it has historically been considered an "insulator" rather than a "semiconductor." Cadmium sulfide is inherently very much less conductive than silicon or germanium; however, when one must work with very thin layers of material, sharply reduced buld conductivity may be an advantage. Considerable work on fabricating thin film field-effect devices of cadmium sulfide and other materials is in progress at RCA laboratories, Princeton, New Jersey, and at Melpar, Inc., Falls Church, Virginia. Melpar, in particular, had a contract for studying "very high temperature" (up to 500 degrees centigrade) active devices for integrated circuits, and has examined some rather exotic materials (for instance, gadolinium sulfide) along with cadmium sulfide; see Thin Film Monotronics - Final Report, 12 March 1964 - 12 March 1965, Melpar, Inc., ASTIA AD 462073.

In the long run, the major advantage of thin-film fabrication techniques over monolithic fabrication techniques if one exists at all may prove to be that a much larger area single substrates can be processed, resulting in larger numbers of interconnected circuits on one substrate. Monolithic techniques are not typically used with single-crystal silicon waffers larger than a couple square inches at present, whereas thin-film techniques are currently being applied to produce passive devices on substrates of much larger area. If thin-film active devices can be made as small, and with as good a production yield, as monolithic field-effect devices are already being made, developments now in progress may eventually allow very large networks of interconnected active and passive thin-film elements on a single substrate. However, thin-film active devices are not yet field-tested, proven components, and definite predictions that they will eventually surpass monolithic active devices are not yet in order.

Thin-film diodes have been made in a "heterojunction" configuration, which means simply a junction between two different semiconducting materials. If the two materials have substantially different band gaps, the one with the higher band gap (that is, the less conductive one) behaves like N-doped material, the other one behaves like P-doped material, and the resulting heterojunction has rectification properties similar to those of a conventional P-N junction diode. Some success has apparently been achieved to date in producing heterojunction diodes of germanium (band gap 0.78 electron-volts) and gallium arsenide (band gap 1.47 electron-volts).

There seem to have been a number of entirely unsuccessful attempts to make bipolar transistors using thin film technology, In the opinion of Dr. R. Pritchard of Stanford University and Motorola Semiconductor, such devices are still about five years away.

1.1.4 Silicon Controlled Switches

The silicon controlled switch (SCS) or "silicon controlled recitifier" (SCR) is not usually classified as a logic element at all, since most existing devices are designed for power handling applications. These devices are composed of four distinct regions of semiconductor material, and they operate somewhat like two bipolar transistors (NPN and PNP) in a tandem arrangement. Once an SCS is caused to conduct in a forward mode by the application of a "turn-on" voltage and a control signal, it continues to conduct (like a gas discharge tube) until the forward voltage is reduced greatly to below the required "sustaining" voltage level. There are even two-terminal four-layer devices (Shockley diodes"), which lack a control lead.. However, an "SCS" must have a control lead, which usually can turn the device "on" but not turn it "off." There is also a newer type of device now available called a "gate-turn-off" (GTO), whose main advantage is that it can be turned "off" as well as "on" from the control lead. With this exception, any such four-layer device must be turned off, once it has been turned on, by cutting off its power somewhere else in the circuit.

SCS devices are included in this discussion because, in a sense, their conductivity properties exhibit hysteresis; hence they intrinsically have "memory" capability. Moreover, it has recently been found possible

to fabricate them in integrated circuit form. SCS's are likely to be somewhat slower than other semiconductor devices; also, they are likely to turn off even more slowly than they turn on, even in the event that they can be turned off via the control lead. Nevertheless, the time has probably come for a serious look at them as active logic devices, at least for integrated circuit memories. After all, an SCS memory would be capable of very fast NDRO readout; the read electronics would merely need to be able to select a resistor in series with the SCS, and then to sense whether the voltage across it was high or low.

1.1.5 Tunnel Diode

The tunnel diode is a two-terminal semiconductor device whose current-versus-voltage response curve is not monotonically increasing, but has a sharp "hump" maximum at some very slight forward bias voltage, superimposed on an otherwise fairly ordinary diode characteristic. Thus its conductivity is much greater for a very small positive bias than for a somewhat larger positive bias; as the bias is further increased in the neighborhood of a volt the curve rises again in normal diode fashion for forward conduction.

This extra hump in the characteristic of a tunnel diode can be explained in terms of quantum mechanics. Charge carriers (electrons or holes) can "tunnel through" a thin potential barrier, such as a very thin P-N junction depletion zone, simply because, for particles the size of charge carriers, the barier is no longer absolute. Instead, given some minute distance into the barrier , there is a small but non-zero probability that the carrier will be found at that position; this probability falls off exponentially as the distance into the barrier increases. If the laws of classical electromechanics still held, of course, the probability would be zero that the particle would penetrate the barrier at all. Now if the barrier is sufficiently thin, for instance the P-N junction depletion zone in a diode in which both the P-region and the N-region are very heavily doped, a given carrier has an appreciable probability of "tunneling" or passing entirely through the barrier to the other side. A very slight positive voltage causes carriers to drift toward the junction, and many pass through it to occupy states on the other side. These states are not available if the forward bias increases, and hence conduction actually falls off until the normal diode forward drop voltage is exceeded, when it rises again.

If the tunnel diode is now placed in series with a resistor, the combination can serve as a memory element capable of very fast switching and also of nondestructive readout. The tunnel diode can be placed in either a normal diode conducting state, or a tunnel conducting state; it will be stable in whichever one it is left, if the voltage supplied to it does not fluctuate and the value of the series resistor was chosen properly to begin with. Graphically, the effect of the resistor can be represented as a "load line" of negative slope, intersecting the hump of the tunnel diode characteristic slightly below its peak and intersecting the forward conduction curve at a lower current value. The two intersections of the load line and the positive resistance portions of the tunnel diode characteristic result in steady-state conditions, and hence are stable points.

1.2 Cryotrons

The foregoing description applies to principle to tunnel diodes in general. Integrated circuits containing tunnel diodes have been fabricated; see "Nonlinear Coupling with Silicon Tunnel Junctions in Integrated Logic," H. C. Josephs, J. T. Maupin, and J. D. Zook, IEE Transactions on Electron Devices, May 1965, pages 237-241. (The authors, are all with Honeywell.) It is reasonable to suppose that tunnel diode matrices, complete with the necessary passive circuit elements, could be fabricated as integrated circuits suitable for use in non-destructive readout memories, or in particular associative memories. Tunnel diodes have the advantages of extremely fast switching and relatively high tolerance to radiation; they have the similar disadvantage that they are two-terminal devices, and the design of logic circuits using tunnel diodes only is difficult.

The cryotron is in principle an extremely promising active element because of its simplicity, inherent batch-fabricability, small size, and "ideal switch" properties. This last term may require a defination; an "ideal switch" is a device capable of making or breaking a zero-resistance contact according to the application of some control signal. A relay is, of course, an ideal switch; but relays are far too slow, expensive, and physically large to be considered for any distributed-logic application such as associative memory active devices.

On paper, cryotrons are very attractive for any application requiring a large, batch-fabricated array of interconnected active elements. The difficulty has proven to be in getting cryotron devices "off paper" into working devices. A very large amount of both government supported and privately supported research effort has been expended to this end in the last 10 years, with scant success; the high point of this work seems to have been the successful fabrication of approximately 120 interconnected, operating cryotrons on a single substrate recently by Arthur D. Little, Inc. At the present time, the company concerned with cryotron work is apparently Texas Instruments, Dallas, Texas; see "Design of a Fully Associative Cryogenic Data Processor", J. P. Pritchard, Jr. and L. D. Wold, IEEE Transactions on Magnetics, March 1965, pages 68-71, and Dielectric Properitis of Thin Insulating Film of Photoresist Material, J. T. Pierce and J. P. Pritchard, Jr., IEEE Transactions on Component Parts, March 1965, pages 8 - 11.

An even more serious strike against cryotronic logic is that its theoretical limiting speed is not far in excess of the actual speed of commercially available semiconductor devices. If cryotrons are ever to compete from now on, it will evidently have to be on a cost basis rather than on a performance basis.

Other types of active element memories are votatile; that is, they are not capable of storing information in the event of loss of system power. Cryotronic memories, on the other hand, will retain their stored information as long as cryogenic temperatures are maintained, with or without the power normally supplied to operate them. This is not to say that cryotronic memories should be considered "non-volatile" either, since power is required to run the refrigeration unit. However, if there is a sufficient reserve of liquid helium in the cryogenic refrigerator, and the memory and the refrigerator are very well insulated, stored information could be retained for hours or possibly even for days. Thus the system can retain

information if the logic power supply fails as long as refrigeration power is not lost, or even if both logic and refrigeration power is lost as long as the period of down-time is relatively short, In order to endow integrated circuit memories with similar properties, it would be necessary to distribute non-volatile memory elements of some type throughout the integrated circuit array; the most likely candidates are ferroelectrc thin film elements, which can be read and written into at power levels compatible with integrated circuit logic.

1.3 Discrete Tunnel Diodes

The operating principles of tunnel diode memories have already been discussed in subsection 1.1.5. If suffices here to remark that discrete tunnel diodes have been a commerically available item for several years, and have been used in experimental non-destructive readout memories. One such memory was operating at Bunker-Ramo Corporation (Then Ramo-Wooldridge Division, TRW, Inc.) about four years ago. Tunnel diodes have been used in varied circuit applications; one relevant computer application is that of providing memory capability within a sense amplifier, which was done several years ago in a mass-produced system in the AN/UYK-1 (TRW-130) computer at Ramo-Wobldridge. For cost reasons, it is unlikely that discrete tunnel diodes would be considered for other than the smallest associative memories; integrated-circuit tunnel diodes would offer a cheaper, batch-fabrication approach which would appear much more attractive in the long run.

2.0 HYSTERESIS ELEMENTS

For the purpose of this document, a hysteresis element is defined as "a device which can be left in one state by a certain type of signal, can be left in an opposite (in some sense) state by a different type of signal, and will remain indefinitely in whichever state it was last left if no further signals are applied to it." Hysteresis elements may be divided into many distinct categories; the way it is done here is certainly not the only plausible way.

This definition of a hysteresis element must be modified for certain compound elements, such as the transfluxor (subsection (2.2.2) or the transpolarizer (subsection 2.5.1)), which are made up of two-state hysteresis materials but may be left in any of n distinct states by various signal combinations, where $n>2$.

2.1 Ferrimagnetic Two-State Elements

A ferrimagnetic or "ferrite" material may be usefully defined for computer engineering purposes as "an insulator which can be magnetically polarized." More precise definitions may be given in terms of solid state physics concepts, in particular that of two opposing polarizable sublattices (see section 2.5 for further comments). Ferrites of commercial importance have one of two general types of crystal structure. One of these is referred to as the "spinel structure" after the mineral spinel, which has the formula $MgAl_2O_4$; this type of ferrite is the one

used in computer memory cores. The other is the "garnet" structure, typlified by yttrium-iron garnet which has the formula $3Y_2O_3 \cdot 5Fe_2O_3$; garnet-structure ferrites are of use in certain microwave devices, but not at the present time in computer memories. A useful reference on spinel-structure ferrites is Solid State Physics, A. J. Dekker, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1957, pages 490-495.

2.1.1 Discrete Toroidal Cores

The most common digital storage element for contemporary central processor memories is undoubtedly the discrete toroidal ferrite core. Such cores can be magnetized to saturation in either of the two possible directions along the circumference. They are customarily operated between these two saturation conditions in a DRO mode, which is less than ideal for most associative memory applications although it has proven very satisfactory for genreal-purpose central processor memories. Nevertheless one DRO associative memory scheme is described in "A Destructive-Readout Associative Memory," Yaohan Chu, IEEE Transactions on Electronic Computers, August 1965, pages 600-605.

It is also possible however, to operate toroidal ferrite cores in a mode using the "center" of the B-H loop, that is, with a core sometimes in a "depolarized" state. In this type of memory, a slight disturbing field applied to a depolarized core results in a fairly large output, because it will cause an excursion around a "minor loop;" but the same field applied to a saturated core results in only a negligible output. For a more complete description, see the section on "minor-loop readout in reference 2. Core memories operating on this principle, generally using two cores per bit, have been studied by a number of computer companies. Also, such memories are produced commercially by at least one vendor, North Electric Company, Galion, Ohio; North Electric has sold them, not for computer applications, but rather for use as "number file" memories for military field command post telephone systems. They feature a read cycle time of from 333 nanoseconds to 2 microseconds, and a write time of 40 microseconds; writing requires a long time because a damped a-c signal is required to leave a core in the depolarized state.

Although early ferrite DRO core memories generally operated with a memory cycle time of from 10 to 20 microseconds, the best contemporary commercially available units operate with about a 1-microsecond cycle time, and 500-nanosecond cycle time memories are in an advanced stage of development. The early memories used relatively large cores having an inside diameter of 50 to 80 mils; contemporary memories use 20-mil or 30-mil cores, and 14-mil core memories will be available shortly.

2.1.2 Approaches to Batch-Fabricated Core Memories

Because of the popularity of the toroidal ferrite core memory, there have been many attempts to lower its cost per bit. Initially, a great deal of study was given to automating the testing of cores and the wiring of core planes, and these tactics have been extraordinarily successful in improving

the cost and reliability of core memories even as cores have gotten radically smaller and more difficult to handle. However, more and more attention is now being given to developing batch-fabricated media which have the essential system and circuit properties of core memories, but which are not composed of individually handled and wired toroids. There are, of course, other possibilities besides continuous ferrite media; some of the ones using permalloy or other metallic (ferromagnetic) material instead of ferrite (ferrimagnetic) material are described in section 2.3.

One of the first such attempts seems to have been a stack of ferrite plates with round holes described in "Ferrite Aperture Plate for Random Acess Memory," Jan A. Rajchman, Proceedings of the IRE, March 1957, pages 325-334. The plate stacks could be wired, using a combination of deposition and discrete wiring techniques, in either a coincident-current or a word-oriented pattern. This scheme, like many others for batch-fabricated core memories, used two storage elements per bit in order to improve noise cancellation and equalize demands on the drive circuits. The work was done at RCA Laboratories, Princeton, New Jersey, which has been actively pursuing the development of memories in this class for at least a decade. This development obviously did not catch on, despite its promise.

Another class of memories, which has received a lot of attention from the telephone industry, is that of "waffle-iron" ferrite sheet memories. These are constructed from a ferrite sheet with a rectangular grid array of "mesas" at close intervals, so that the total effect is like that of the metal plate on half of a waffle baker. Wires run in two directions, along the grooves between rows of mesas and columns of mesas. To close the flux paths, a thin sheet or thin film of permalloy on a substrate is placed on top of the array under enough pressure that it makes physical contact with each mesa; a "core" is then the cylinder formed by the walls of two adjacent mesas, the floor of the groove separating them, and the permalloy "roof." For details see "The Cubic Waffle-Iron Memory," A.H. Bobeck, 1963 Proceedings of the Intermag Conference (an IEEE publication), pages 3-2-1 through 3-2-6, which is an account of work at Bell Labroatories, Murray Hill, New Jersey; also "latest Developments in Ferrite Memories," Computer Design, August 1965, pages 42-48, which includes an account of work at International Telephone and Telegraph (England) on pages 46-47.

A relatively recent major development in this class of memories is also briefly described in the above Computer Design article, and in much more detail in "Laminated Ferrite Memory," R. Shahbender, C. Wentworth, K. Li, S. Hotchkiss, and J.A. Rajchman, AFIPS Conference Proceedings - Volume 24 - 1963 Fall Joint Computer Conference, pages 77-90; Spartan Books, Inc., Baltimore, Maryland. A typical laminated ferrite structure would use three thin sheets of ferrite; wiring is deposited on two of the sheet surfaces, in such a way that when the three sheets are stacked the interior of the sandwich contains two arrays of parallel printed wires, oriented orthogonally to each other, and

separated by a thickness of ferrite which insulates the wires in one array from those in the other. Each wire intersection point defines a bit position, although two positions per bit would be used in practice. This sandwich is sintered under pressure, and holes are drilled through it to make connections to the wiring. The resulting plane is suitable for high-speed, word-oriented operation. This development is also due to RCA Laboratories.

These three classes of memory configurations are fairly typical, of what can be expected, but the list could be greatly extended. Each organization developing such a memory makes impressive claims for it, but it is almost impossible to sort out those which will catch on in a big way from those which will not. None are yet in use in commercial computer systems to the best of my knowledge, although waffle-iron memories are probably in use to a limited extent in telephone switching installations by now. There are certain generalizations about such batch-fabricated analogs of core memories which probably are valid for most present and future developments:

(1) They always have a closed-flux geometry; each element is efectively a "toroid" for magnetic flux path considerations.

(2) They tend either to be word-oriented or else to use the new 2-1/2D organization, for the simple reason that it is easy to have two intersecting wires at a bit position and hard to have more than two, and two wires per bit position suffice for word-oriented operation but not for coincident-current operation. Also, word-oriented operation is usually somewhat faster.

(3) The goal of both geometrical design and circuit design (i.e. the use of partial switching) for such memories is to make the effective size of a closed-flux ferrite "core," no matter how such a "core" results from the fabrication process, as small as possible in order to obtain increased speed.

(4) The practicality of such memories cannot readily be determined until they have not only been proven in principle, but also put into pilot production, since questions of manufacturing cost and yield are at the root of their claimed superiority over discrete toroidal core memories and over each other.

(5) Claimed speeds are hard to assess, because often a well-tuned laboratory prototype may safely be operated at a much higher speed than a production model available at a reasonable cost per bit.

(6) With the exception of the waffle-iron memory, developments in this category are generally quite dissimilar from one company to another - in contrast to plated-wire memories (subsection 2.3.2), which have shown convergence instead of divergence of individual lines of development at different companies.

(7) Probably most important for associative memory purposes, batch-fabricated analogs of core memories are virtually always intended solely for DRO operation; geometrical considerations have thus far ruled out NDRO operation (except perhaps for the partial switching mode used by North Electric (see subsection 2. 1. 1)), and so a DRO associative memory organizational scheme such as that of Chu (see also subsection 2. 1. 1) would have to be considered given this choice of memory element.

Claimed speeds are as follows: 1. 5 microsecond rise time (meaning probably a total cycle time of several microseconds) for the ferrite apertured plate, 200-400 nanoseconds total cycle time for Bell Laboratories' waffle-iron memory, 500 nanoseconds total cycle time for British ITT's waffle-iron memory, and as low as 200 nanoseconds total cycle time for the laminated ferrite memory. It seems safe to predict that whatever scheme of this class works out best from a manufacturing point of view will find considerable acceptance for future DRO memories; however, in the absence of a well-developed NDRO scheme these memories will not compete too well with plated-wire memories for use in associative systems, or for that matter for use in militarized or industrial control system general-purpose computers.

2. 2 Ferrimagnetic Multi-State Elements

Devices in this category have generally been described under the name of "multiaperature cores," but this term implies only that the device is a chunk of ferrite with more than one hole through it, which nevertheless is used to store only one bit of information. The manner in which the multiple holes affect the switching properties of the core varies considerably, depending on the type of device. There are essentially two major categories of multiaperture cores, one of which features true NDRO readout, and the other of which features, "pseudo-non-destructive" (PNDRO) readout (see subsection 2. 2. 2).

2. 2. 1 BIAX and MicroBIAX

A BIAX is a rectangular solid ferrite magnetic storage element with two cylindrical holes through it at which are "skewed," that is, they are at right angles but do not intersect inside the core. One of these holes, the "storage hole," is typically somewhat larger in inner diameter than the other. Information may be stored in this hole by any of the techniques suitable for a toroidal ferrite core memory. The other hole is the "interrogate hole."

The operating principle for interrogation is that a fairly strong pulse through the interrogate hole does not, by reason of the geometry of the core, actually switch flux around the storage hole unless this pulse is of extremely great magnitude. Thus, readout can be performed in a BIAX memory by applying a pulse in the interrogate line which will tend to rotate the flux in the volume of the core between the two holes. When the pulse is released, this flux returns to the condition it was in before the interrogate pulse was applied. The net flux change can be observed as a brief a-c signal by a sense wire through the storage hole; it will be in one direction for one

sense of magnetization around the storage hole and in the opposite direction for the other sense. A BIAX memory, particularly a microBIAX memory, is similar in system design properties to a plated wire memory, even though the fundamental physical interactions within a storage element are quite dissimilar. Thus, it is virtually true that any associative memory system design which is sound for a microBIAX memory would be sound for a plated-wire memory and conversely.

2.2.2 Transfluxors

The term "transfluxor" here means essentially any "pseudo-non-destructive readout" (PNDRO) multiaperture core; there are many types of these cores on the market, not all of which are sold as "transfluxors." The name "transfluxor" has been applied principally to two-apertrue devices; vendors producing cores with more than two apertures have, probably for marketing reasons, chosen to call them by other names, such as MALE (Multi-Aperture Logic Element, Goodyear Aerospace Corporation, Akron, Ohio) or MAD (Multi-Aperture Device, AMP Incorporated, Harrisburg, Pennsylvania. The basic operating principles of the transfluxor are well presented in the following reference: Digital Computer and Control Engineering, R.S. Ledley, McGraw-Hill Book Company, New York, 1964; pages 706-711 concerns transfluxors.

Several different flux configurations are possible in a transfluxor. There are at least two holes; the "storage hole," and the "readout hole." If the flux around the readout hole can be readily switched, the transfluxor is said to be in an "unblocked" state; otherwise, it is said to be in a "blocked" state. Transfluxor memories do not operate in a truly NDRO mode; the readout process partially destroys the original flux configuration in "unblocked" cores and does not affect "blocked" cores, but a master "restore" pulse for the entire memory can be used immediately after readout to re-establish the original unblocked flux configuration in every core where it was destroyed. Thus transfluxor memories are like DRO core memories in that flux is switched to a new stable configuration by the readout process, whereas they are like NDRO core memories in the information in the cores is not destroyed by the reading operation even though the flux configuration changes. Conventionally, the blocked state signifies a 0 state and the unblocked state a 1 state; the disturbed unblocked state, which results in a core after the original unblocked state is destroyed by the readout process, is usually denoted as the 1' state (read "one-prime").

Thus, the readout pulse does not affect cores in the 0 state, but changes selected cores in the 1 state to the 1' state. The restore pulse likewise does not affect cores in the 0 state; it does not affect cores in the 1 state either, since it merely tends to saturate the flux around the readout hole further in the 1 direction; but it will return any core in the memory in the 1' state to the 1 state. The illustration on page 708 of the aforementioned book by Ledley shows the 0 and 1 states; the 1' state is similar to 1 state except that the flux direction around the readout (smaller) hole is reversed, and the flux around the storage (larger) hole is somewhat disarranged. There are, incidentally, six possible states in all for a transfluxor, since the same three states can be equally well defined with the flux around the storage hole reversed in direction.

Associative memories based on two-hole transfluxors would normally use two per bit. Such memories have been developed by Scope, Inc., Falls Church, Virginia, and by RCA Laboratories, Princeton, New Jersey; see "Transfluxor Content-Addressable Memory," A. D. Robbi and R. Ricci, 1964 Proceedings of the Intermag Conference (an IEEE publication), pages 8-3-1 through 8-3-7. A memory developed by Goodyear in Akron uses one three-hole MALE transfluxor to do the work of two two-hole transfluxors, and so it has one core per bit; the method by which this is accomplished is also described by Robbi and Ricci. The use of signal cancellation techniques for logic at the bit level is natural with transfluxors, since their output signals are relatively large - which is a consequence of the flux switching which takes place during readout, and hence comprises an advantage of PNDRO devices over true NDRO devices. As already mentioned, AMP Inc., manufacures a line of "MAD" cores; these are essentially five-hole transfluxors, and are intended for use as magnetic logic elements. However, they are quite slow, and no attempt appears to have been made to base an associative memory design on them.

The name "transfluxor" generally implies a ferrite core, but a permalloy sheet "transfluxor" memory has also been developed by RCA (see subsection 2. 3. 3). A feroelectric device whose operation is rather directly analogous to that of the transfluxor is the "transpolarizer," discussed in subsection 2. 5. 1.

2. 3 Ferromagnetic Thin Film Elements

For the purpose of this discussion, a ferromagnetic material is "a magnetically polarizable conductor." Here, as for ferrimagnetic materials, the definition in terms of solid-state physics concepts is considerably more subtle. The most useful way to categorize existing ferromagnetic thin film memory devices is simply according to whether each bit of storage operates on an open-flux principle, as do flat thin-film memories made by Univac, Fabri-Tek, Philco-Aeronutronic, and other companies; or on a closed-flux principle, as do the plated wire memories of Bell Laboratories, Honeywell and Toko, the woven-screen memory of Bunker-Ramo Corporation, and the "rod" memory developed by National Cash Register Corporation.

The difference between these two types of thin magnetic films is that open-flux films have no return flux path except through air or the substrate, while closed flux films have such a path through magnetically polarizable material. However, for the purpose of systems logic description, there is essentially no difference between open flux and closed flux elements. Both can be operated in either a DRO or an NDRO mode, although the geometry of the element may change according to the desired mode of operation. Although open-flux elements can be more easily batch-fabricated in large numbers, because the fabrication can be entirely a matter of masking processes on a substrate, they have not been very successful in the commercial market, mainly because of extremely low readout signal capabilities and "creep" (unwanted partial switching) problems. Commercial devices with memory cycle times down to 300 nanoseconds are available from Burroughs and Fabri-Tek, and possibly from other vendors.

2.3.1 Open-Flux Thin Film Elements

Open-flux thin-film elements, or "flat film" elements, have been a subject of widespread and intensive investigation by the computer industry for probably almost a decade. Virtually all of the dozen or so largest computer manufacturers have tried their hand at development of flat-film memories, as have several major vendors who supply components to the computer industry. Although these efforts have by no means ended in the almost complete failure evident in the similarly large effort on cryogenic memories, there have been no really solid successes either, and to this day very few flat-film memories are in use in operating computer systems. Although it was freely predicted some years back that flat-film memories would rapidly displace core memories, there is still not a single commercially available computer system featuring a flat-film main memory. To be sure, flat-film scratchpad memories have found some commercial acceptance, and there are a few prototype flat-film memory aerospace computers.

A flat-film memory plane consists of an insulating substrate, with areas of very thin permalloy film, criss-crossed by flat sense and drive leads. There may be one film spot, or two, per bit storage position. The storage element geometries and system operating modes of these memories vary too much to summarize here. All flat-film memories, however, qualify as "batch-fabricated" memories. They could have a considerable cost advantage over core memories in the future for this reason, but at the present time the manufacturing yield for flat-film planes is poor enough and the circuit requirements for complete flat-film memory systems are severe enough that they are much more expensive than core memories. Where flat-film memories have been used in computers in preference to core memories, the reason has generally been either their greater speed, or their compactness and potentially superior tolerance of unfavorable thermal and vibration environments.

Flat-film random-access memories and subsystems are offered commercially by Burroughs Electronic Components Division, Plainfield, New Jersey, and by Fabri-Tek, Edina, Minnesota; Philco-Aeronutronic, Newport Beach, California, at one time was also pursuing this market. Flat-film domain-wall shift registers (see subsection 4.2.2) are also available commercially from at least two sources. Univac, St. Paul, Minnesota, and Blue Bell, Pennsylvania, has been the most active in flat-film work of any of the large computer manufacturers, followed by Burroughs, Paoli, Pennsylvania. Other significant work has been done at universities, telephone company research laboratories, and government facilities both in this country and elsewhere; among the latter, the Centre Energie Atomique de Grenoble (CENG), and the Centre National de la Recherche Scientifique, (CNRS), both French government facilities located at Grenoble, France, have been particularly active.

The two computer systems making the most significant use of flat-film memories to date have been the Univac 1107, and the Burroughs D-825 and its various successors. Each of these systems has been produced in sufficient quantity (probably between one and three dozen copies) to establish the type of memory used in each as a proven commercial component. The 1107 has a 600-nanosecond cycle-time memory of 128 36-bit words; these are used both to store a few frequently used instructions, and to function as machine-addressed control registers such as index registers. The D-825 has a smaller memory, which is used only for control registers.

The films used in flat-film memories are typically very much thinner than those used in closed-flux (plated-wire or "round-film") memories, discussed in the next subsection; typical thicknesses are 1000 angstroms (10^{-5} cm) for flat films, and 10,000 angstroms (10^{-4} cm) for round films. CENG has investigated much thinner flat films also, down to 350 angstroms; several complete memories have been fabricated and sold to Frence industry, the largest being a 5-megacycle unit having 512 48-bit words. Some detailed comparisons are made in a recent article, "Comparison of Magnetic Behavior of Cylindrical and Flat Films from Kerr Effect Probe Measurements," D. B. Dove, T. R. Long, and J. E. Schwenker, IEEE Transactions on Magnetics, September 1965, pages 180-185. A direct consequence of this disparity in thickness is an equally great disparity of the output signal voltages normally encountered - 1.5 to 3 millivolts for flat-film memories, 7 to 25 millivolts for round-film memories. The drive currents used in round-film memories are indeed greater also, but not by nearly as large a factor. It is a fair question as to whether flat-film memories can be used in standard main-memory configurations for central processors (32,768 48-bit words is rather standard as of this writing) without a substantial improvement in the obtainable output signal amplitude. However, the very thin films are reportedly less prone to "creep," which is a cardinal advantage.

Most of the flat-film memories in existance are intended for DRO operation. Univac (St. Paul) and Aeronutronic have been active in the development of NDRO flat-film memories, with claimed readout rates of 10 and 5 megacycles respectively. Work now in progress at CNRS could lead to an NDRO flat-film memory based on a coupled-film "sandwich;" see "A Coupling Phenonemon Between the Magnetization of Two Ferromagnetic Films Separated by a Thin Metallic Film - Application to Magnetic Memories," J. C. Bruyere, O. Massenet, R. Montmory, and L. Neel, IEEE Transactions on Magnetics, March 1965, pages 10-12. Similar studies are in progress under a NASA contract at the Honeywell Corporate Research Center, Hopkins, Minnesota, with respect to coupled concentric round films.

References containing extensive bibliographies on flat-film memories are: "Storage Systems - Present Status and Anticipated Development," H. P. Louis and W. L. Shevel, Jr. IEEE Transactions on Magnetics, September 1965, pages 206-211; "High Density Magnetic Film Memory Techniques," T. S. Crowther, pages 5-7-1 through 5-7-6, 1964 Proceedings of the Intermag Conference (an IEEE publication), and "Amplifier and Driver Circuits for Thin Film Memories with 15 Nanoseconds Read Cycle Time," D. Seitzer, IEEE Transactions on Electronic

Computers, December 1964, pages 722-729. This last paper certainly indicates what can be done on a laboratory basis if one really tries.

2.3.2 Closed-Flux Thin Film Elements

Closed-flux memory elements are more expensive to fabricate than open-flux elements, because of the necessity of plating magnetic material on some sort of cylindrical conducting substrate and putting another conductor orthogonally to it. Nevertheless, clased-flux memories have proved to be more easily reduced to commercial practices, probably in large part because the readout signal for typical drive currents ranges from 7 to 25 millivolts - compared to 1.5 to 3 millivolts for open-flux devices. The only closed-flux memory now in widespread commercial use is that of National Cash Register, Dayton, Ohio and Hawthorne, California, it operates as the main memory in their reengineered "rod memory NCR 315" computer with a cycle time of 800 nanoseconds. However, plated-wire memories are likely to be very common in computers in the near future.

Plated wire memories ultimately will probably be useful at readout times (NDRO) as fast as 100 nonoseconds and write times of 300, since speeds of this general order have been obtained by various techniques in the laboratory; however, the speeds announced in production-model computers (such as NCR's) will be slower than this for a period of time. There are three principal read-write operating modes which may be used with a typical plated wire memory element, which may be called DRO, NDRO-1, and NDRO-2. In DRO mode, the interrogate wire carries a current pulse of height I_w', which results in a field of amount H_w' in the plated-wire element where H_w' suffices to rotate the magnetization vector through 90 degrees and thus to destroy any previously stored information, which can nevertheless be sensed at the time of destruction by a sense amplifier; a small current I_d' is then put through the sense wire in order to rewrite or to write new information, and it produces a field H_d' sufficient to cause the magnetization vector to fall back to whichever one of the two available and opposite no-field orientations is specified by the choice of the direction of I_d'. In the NDRO-1 mode, writing is accomplished in the same manner as in the DRO mode; but reading is accomplished nondestructively by the use of a current pulse of smaller height I_w, which produces a field H_w which is insufficient to rotate the magnetization vector through 90 degrees but is sufficient to produce an output signal capable of being picked up by the sense amplifier. In the NDRO-2 mode reading is accomplished nondestructively in the same manner as the NDRO-1 mode; writing is still accomplished using the same current pluse height I_w suitable for nondestructive reading, and applying a much larger sense wire current I_d whose resulting field H_d is enough to push the magnetization vector "up" through 90 degrees and "down on the other side" if the information bit stored in the plated-wire element is to be reversed. Very rough numeric ranges for these various currents are: I_w', 700 to 1000 milliamperes; I_d' 7 to 20 milliamperes; I_w, 400 to 600 milliamperes; I_d, 60 to 100 milliamperes. High-voltage switching transistors are now available that are capable of handling these currents with the necessary rise times and driving the characteristic

impedances associated with the application, although the use of an NPN-PNP driver transistor pair in "push-pull" may be necessary where the voltage required to develop the current exceeds about 50 volts.

The requirements which each of these three operating modes places on the bit-to-bit uniformity of a plated-wire memory plane, increase in severity in the order indicated: DRO, NDRO-1, NDRO-2. The NDRO-2 mode is often considered to be the least expensive in terms of circuits, but the situation si perhaps not quite that simple. To be sure, the NDRO-1 mode requires two different levels of drive current in the interrogate wire, one for readout and one for writing, and the writing level is quite high; however, there is something saved in the driver which must be coupled to the sense wire, since the required output is now so small that this driver essentially becomes a glorified logic circuit. Moreover, it is possible that some arrangement can be worked out using complementary drive transistors, such that only the NPN functions during reading, and both it and the PNP function during writing, to conveniently cope with the requirement of two distinct current pulse levels in the drive wire. In any case, it is obviously extremely desirable to consider carefully any circuit configuration, even if it is more expensive, which promises to relax the bit-to-bit uniformity requirements in plated-wire memories; these have been a problem with all approaches, and are a corresponding greater problem in those approached in which the manufacture of memory planes is done with a very high degree of batch-fabrication.

Several types of closed-flux memories have been seriously investigated by various computer development organizations; as these various development programs progressed, they have in most cases tended to "converge" toward similar geometries and operating philosophies. The first of these to fully succeed is NCR 800-nanosecond rod memory, which began with a glass substrate but went into production with a beryllium-copper conducting substrate. See "A Five Megacycle DRO Thin-Film Rod Memory," D. A. Meiar, 1963 Proceedings of the Intermag Conference (an IEEE publication), pages 9-4-1 through 9-4-11.

The second oldest memory of this type, in terms of the point of time of when it was actually started, is probably the woven screen memory developed by Bunker-Ramo Corporation (formerly Ramo-Wooldridge Division of TRW, Inc.) Canoga Park, California. This memory is designed principally for coincident current destructive readout, although other modes of operation are possible in principle. A reference is "A Technical Note on the Application of Wearing to the Batch Fabrication of Electronic Components and Subsystems," J. Davis, Proceedings of the National Symposium on the Impact of Batch-Fabrication on Future Computers (an IEEE publication), 6-8 April 1965, pages 67-80.

A third development is the "screen door" memory at IBM," there are some brief comments made about this memory in a recent paper, "Storage-Systems - Present Status and Anticipated Development," H. T. Louis and W. L. Shevel Jr., IEEE Transactions on Magnetics, September 1965, pages 206-211. This article is worthwhile reading for anyone interested in computer memories.

Another well known development is by Toko, Inc., Tokyo, Japan. For information on this Toko's memory see "Woven Thin-Film Wire Memories," H. Maeda and A. Matsushita, IEEE Transactions on Magnetics, March 1965, pp. 13-17. Toko memories are being supplied to the Librascope Division of General Precision Equipment, Inc., Glendale, California, for use as associative memories and for other purposes. A paper on this memory by Librascope authors is "Some Considerations in the Design of Plated-Wire Memory Systems," M. Bienhoff, J. Camorate, and M. Sherman, Proceedings of the National Symposium on the Impact of Batch Fabrication on Future Computers (an IEEE publication), 6-8 April 1965, pages 88-102.

Bell Laboratories, Murray Hill, New Jersey, has been a leader in this field for years. A short and informative article summarizing the Bell Labs work is "Plated Wire Magnetic Film Memories," U. F. Gianola, Bell Laboratories Record December 1964, pages 408-411.

Taken as a class plated-wire memories appear to be the single most plausible development of any now on the horizon which are suitable for the full range of computer memories in which discrete toroid cores are now used. It is probably too early to say just which fabrication approach to plated wire will be the most successful in the long run, but it is not too early to say that at least one of them will be very successful. Although plated wire memories have not yet reached the degree of development at which their characteristics are well understood and controlled as those of core memories, they are far enough along now that they can almost surely reach that point within a short enough time to make them of immediate interest to product-line organizations. As a class, closed-flux thin-film or "round-film" memories have two very great advantages over flat-film memories. First, it is possible to obtain the same characteristics with much thicker films in the closed-flux case without the demagnetization problem. Because of that and the better coupling resulting from the closed-flux configuration the readout signal is very much stronger, frequently ten times as strong as from a flat film memory. In the second place, the fact that the flux path is closed in the "easy" direction makes the memory elements relatively insensitive to outside stray fields, and thus plated wire memories would be somewhat better suited for operation in normal environments. (See "Comparison of Magnetic Behavior of Cylindrical and Flat Films from Kerr Effect Probe Measurements," D. B. Dove, T. R. Long, and J. E. Schmenker, IEEE Transactions on Magnetics, September 1965, pages 180-185.) The vastly increased readout signal strength implies that round-film memories will attain, in the long run, both greater reliability and higher operating speeds than flat-film memories, with the further inference that larger size memory modules are more likely to be economically attractive. It is not unreasonable to anticipate operating NDRO rates of 10 megacycles and write rates of 3 to 5 megacycles within the time frame of this study (that is, by 1975) in memories in mass production.

Plated wire and other closed-flux memories can be used for associative memories in at least two distinct ways. First, they can be operated as a simple nondestructive readout memory with associative logic functions implemented by external logic per word. Secondly it may be possible to use signal cancellation techniques, originally developed for discrete cores and flat-film memories to perform a certain amount of "local" logic right at the physical location of a stored bit of information.

2.3.3 Metal Cores

The permalloy tape-wound core has been in use for many years as a circuit element, although it has been used rarely in computer memories. Such a core is made by starting with a "bobbin" of stainless steel, nylon, or ceramic material, and winding on that bobbin several turns of very thin permalloy metal tape, typically between 2 and 20 turns. The thickness of the tape used ranges down to 1/8 mil. The purpose of manufacturing permalloy metal cores by this winding method, instead of by the more straightforward one of simply forming them out of solid permalloy, is to cut down on eddy current loses - which would otherwise be severe, because permalloy is a metallic conductor. Once the tape has been wound, the core is typically filled with an encapsulant and sealed. The resulting cores may each cost 50 cents or more, even in some quantity, but their properties are otherwise excellent for many computer applications; they are fast, uniform and able to withstand unfavorable conditions.

Unfortunately, since the fabrication process in making tape-wound cores is very complex, substantial price cuts are improbable; for this reason, it is not likely they will ever be used in main-frame computer memories, although they would be reasonable components for scratchpad memories to operate at speeds up to several megacycles. However, equally fast alternatives for this application are in sight which will not be nearly as expensive per bit; for instance integrated circuit memories, laminated ferrite memories, and plated-wire memories.

The desirable properties in a metal core are somewhat wider temperature margins than ordinary ferrite devices, great physical strength, very good hysteresis loop squareness in many permalloy compositions, and extremely fast switching characteristics. The question then naturally arises, if there is not some batch-fabricated form of the permalloy tape-wound core which preserves these advantages, and yet achieves a drastic reduction of the cost per bit. The closest thing yet to such a device is a mass memory element under development by Laboratory for Electronics, Inc. (LFE), Boston, Massachusetts. A short description of this memory is "System and Fabrication Techniques for a Solid-State Random-Access Mass Memory", H. W. Fuller, T. L. McCormack, and C. P. Battarel, IEEE Translations on Magnetics, March 1965, pp. 21-25. LFE's memory basically consists of a substrate on which is bonded a thin sheet of permalloy; such sheets are available with excellent uniformity of properties. By photoetching techniques, this permalloy is then formed into many small toroids, each having the shape of a thin washer. The wiring is then performed by repeated photoetching and electrodeposition, until every one of these permalloy washers has three wires passing through it. Writing in this memory is by conventional coincident-current techniques. Readout, however, uses a highly unconventional "beat-frequency" continuous-wave technique, which operates as follows: The X-direction drive wire carries a sine wave signal of frequency ω_1, and the Y-direction drive wire carries a sine wave signal of frequency ω_2. In one entire plane of cores, exactly one core will have both frequencies passing through it on the dirve wires linking it. The core functions as a

"non-linear mixing element" and produces a complicated output wave form, of which a major component is at the frequency $\omega_3 = \omega_1 + \omega_2$. The sense amplifiers are equipped with narrow-pass filters, set to accept only frequencies close to ω_3; the phase of the ω_3 output signal then varies by 180 degrees according to whether the core was in the 1 state or 0 state. This readout process is entirely non-destructive.

Although LFE plans to produce planes of size 256x256, they were in rather early stages of development when the cited paper was written, and they were contemplating only small test planes of the size 4x4 and 16x16. The scheme is quite novel and has apparently not yet been demonstrated in any configuration even close to a complete memory. Nevertheless, LFE is confident that the ultimate cost per bit should be quite low, in view of the extent to which planes are batch-fabricated and the expected high yield.

LFE's bonded permalloy sheet approach to producing planes with washer-shaped cores on them is by no means the only one possible. Honeywell Corporate Research, Hopkins, Minnesota, has developed a method of using "electroless" chemical deposition techniques to produce laminated permalloy structures, which would have lower eddy current losses than an equivalent amount of solid permalloy. It might therefore be possible to produce planes of the same basic geometry by an entirely different fabrication method, in which layers of permalloy with interleaving insulating layers were built up on a substrate purely by means of chemical deposition. Of course, LFE's reason for choosing the permalloy sheet method appears to have been mainly the possibility of obtaining very good bit-to-bit uniformity that way, since the sheet itself if quite uniform; adequately small tolerances on bit-to-bit variation of electrical characteristics might be more difficult to achieve by the chemical deposition route.

Despite the great speed at which permalloy elements can be switched, there appears to be no reason to believe that the beat-frequency NDRO scheme will lead to really high-speed operation. However, it is quite likely that a more conventional readout scheme, used with smaller planes, could result in a very competitive form of DRO, batch-fabricated "core" memory; LFE has apparently not been as interested in this type of configuration as in mass memories. Such a memory, is organized in a "word-oriented" or "linear-select" configuration, might be used as an associative memory according to the scheme proposed at Control Data Corporation, Rockville, Maryland; See "A Destructive Readout Associative Memory," Yaohan Chu, IEEE Transactions on Electronic Computers, August 1965, pp. 600-605.

Permalloy sheet transfluxors have been investigated extensively by RCA Laboratories, Princeton, New Jersey; an experimental memory based on three-hole permalloy sheet transfluxors has in fact been constructed as recounted in "Design and Operating Characteristics of a High-Bit Density Permalloy Sheet Transfluxor Memory Stack," G.R. Briggs, and J.W. Tuska, 1963 Proceedings of the Intermag Conference (an IEEE publication), pages 3-4-1 through 3-4-8. The operating principles of two-hole transfluxors are discussed in subsection 2.2.2; regarding three-hole devices,

see the paper just cited or else the Robbi and Ricci paper cited in subsection 2.2.2.

2.4 Electrostatic Devices

The devices considered in this section make use of an information storage principle which is fundamentally capacitive; that is, they store a bit of information as a charge or the absence of a charge. The capacitive elements used in these devices generally do not exhibit any significant hysteresis effect unless semiconductor devices are used in series with them, in such a way that their normal tendency to spontaneously discharge is held in check.

Electrostatic storage devices are quite out of fashion at present , but is appropriate to include them in a survey like this for two reasons: (1) The techniques applied were made to work pretty well when digital computer technology was relatively crude, and they have had very little attention since; therefore, this area is one where some technological surprises might be forthcoming. (2) Although capacitors and semiconductors may be implemented either as continuous media or as arrays of discrete devices, individual elements are very compatible with modern batch-fabrication techniques. It is therefore not impossible that there will be an abrupt revival of one or more of these early concepts in a very different form, and that this form will be useful for the disign of associative memories and processors.

2.4.1 Williams Tube

The Williams tube dates from 1949. It has been very much out of fashion for about the last ten years, and almost a whole generation of computer designers has entered the field who have forgotten that a) it was the first random-access memory device to receive any degree of widespread application and to achieve any real degree of practicality, and b) it was also the first batch-fabricated random-access memory device. In fact, Williams tube memories can justly be considered "batch-fabricated" to a much greater degree than magnetic toroidal core memories. The Williams tube had various unpleasant drawbacks, and these led to its eclipse; but it was still in the running as late as the time when the IBM 704 was being designed. After all, the core memory represented an apparent retreat from a continuous, batch-fabricated medium to a discrete array of individually fabricated bits of storage; this retreat proved to be beneficial, but during the early 1950's it was not an obvious move.

The basic principle of the Williams tube memory is as follows: A cathode-ray tube with a phosphor face, of the ordinary variety used in oscilloscopes and television receivers, can create on its face a "potential well" or small spot of positive potential after its electron beam bombards a spot, if the beam is operated at a sufficient voltage (about 3000 volts). This potential well continues to exist for at least a few milliseconds, and therefore is potentially a means of storing information. Most recording schemes used involved two adjacent potential wells which interfere with each other in a predictable manner. The face of the storage tube holds an array of dots which range from 16x16 to, in some cases, as much as 64x64.

Readout is strictly destructive. When a potential well already in existence is again bombarded, a pulse can be detected by a conductive plate, which is placed over the face of the tube; but after that the bit just read must be regenerated immediately. This plate is connected to a sense amplifier. One tube, plate, and sense amplifier corresponds logically to a core plane, sense winding, and sense amplifier in a coincident current core memory.

Depending on the nature of the electronics, the coding used to represent a 1 state or a 0 state may be of various potential well configurations. Williams tube memories may be operated at read-write cycle times of from 10 microseconds down to 2 microseconds; 2 microseconds apparently was attained only under laboratory conditions, whereas 10 microseconds was attained in operating computers.

The use of cathode-ray tubes is at once the great strength and great weakness of Williams tube memories. The drawback of volatility, while serious, is not fatal for non-real time, ground-based computer systems, since the memory can be made to appear non-volatile to the computer user by means of periodic regeneration of all of the information in the memory. (However, a Williams tube memory would be highly inappropriate for a real-time control computer.) Not only are cathode-ray tubes continuous media, but they come ready-equipped with a built-in two-dimensional selection mechanism, electron beam deflection, which is fast and easy to control although it has calibration problems.

Now it is true that, over a period of time, the array of bits may shrink slightly, or cease to be square, as television pictures may in older tubes. Thus the binary code for selection of some given bit may come to select a different physical point on the tube surface as the tube ages. However this decalibration process is gradual, and the information is periodically regenerated over the entire array of bits; hence this effect is not serious in a Williams tube memory, although it would be serious in a non-volatile memory. Nevertheless, there could come a time when the bit pattern shrinks so much that adjacent bit positions start to mutually interfere in the wrong way, and unreliable operation would then ensue.

The question then is, "Is there any conceivable analog of the Williams tube which retains its basic advantages and overcomes its limitations?" At present the use of beam deflection techniques is only practical with electron beams. There are to be sure, various schemes for deflecting light beams, but unfortunately light beams still cannot be deflected through a sufficient angular amplitude to select enough discrete points in a storage plane, except by schemes which are either prohibitively costly or else use mechanical moving parts. And, although electron beams can readily be deflected, the deflection process is probably not precise enough, and the phenonema which can be controlled by the position of an electron beam are not as useful as those which can be controlled by the position of a light beam.

Work is currently in progress on the control of laser beams by solid-state deflection techniques; see "Laser Displays," V. J. Fauler, Data Systems Design, September 1964, pages 50-51. If this work is successful, the ability to build a good laser-beam oscilloscope would directly imply the ability to build an electro-optic memory based on photoelectric material controlled by the coincidence of a laser-beam spot and and applied voltage.

Ferrotron and silverplate memories (see section 2.6) could be subject to control by a laser beam, or even by two laser beams at different frequencies. These memories are examples of a general class of memory based on a "sandwich" of photoconductive and hysteresis materials.

It is even possible that techniques can be devised to render ferroelectric or other information storage elements directly controllable by a deflected electron beam. If very stable and accurate electron-beam deflection devices were available, such a development could cause quite a sensation in the memory business.

A technical reference on the Williams tube is a short article by the inventor, Fredrick T. Williams, on pages 12-34 to 12-41 of the Computer Handbook, edited by H. D. Huskey and G. A. Korn, McGraw-Hill Book Company, New York, 1962.

2.4.2 Diode-Capacitors Pairs

The diode-capacitor or "dicap" memory is probably the first one ever constructed which attained speeds of several megacycles in random-access operation. In early computers build by the National Bureau of Standards, Washington, D. C., various configurations of dicap memory were quite popular. The most recent of these computers, PILOT, was a machine which would have been approximately in the class of the Control Data 3600, had it ever been completed with the full range of equipment contemplated. It was originally to have had a dicap memory of 32,768 72-bit words, with a cycle time of one microsecond; but for budgetary reasons this configuration was cut bact to a 256-word dicap scratchpad memory plus a large core memory. The PILOT computer was actually a multi-processor, and there were other, smaller dicap memories included elsewhere in the system.

The only commercial usage of the dicap memory seems to have been in the British EMIDEC 2400 computer, a machine falling roughly in the class of the IBM 709 or Philco 2000-210 and contemporary with them. Only a relatively small number of EMIDEC 2400 systems were produced. The dicap memory was a scratchpad containing 64 36-bit words, and had a cycle time of 4-1/2 microseconds; the main memory was a 10 microsecond core memory.

The dicap memory uses no memory elements having physical hysteresis properties, and consequently all operations within such a memory can be carried out at the speed associated with semiconductor circuits. Despite the relatively crude state of semiconductor components 7 or 8 years ago, the speed obtained in some of these memories was outstanding even by contemporary standards.

Like the Williams tube memory, the dicap memory requires periodic regeneration of the stored information. However, there are ceratin applications, such as that of a lookahead memory in a large computer, where this attribute might not interfere with the usefulness of the memory. There are others, such as that of a main-frame spaceborne associative memory, where it would. There is still reason to consider a dicap memory for a

small ultrahigh-speed spaceborne associative memory to be used in conjunction with a much larger, non-volatile associative memory; it would not be absolutely necessary, in that situation, that the small one be non-volatile.

The basic circuit of the dicap memory consists of two semiconductor diodes in series, which normally are both kept strongly back-biased. At the point where the one diode is tied to the other, a capacitor is connected which is then grounded through a resistor. Between the capacitor and the resistor is the terminal used for both reading and writing. Selection of a capacitor is accomplished by grounding both diodes; then, if there is any voltage stored on the capacitor, it will spontaneously discharge through the resistor, with one diode or the other conducting depending upon the sign of the voltage which was previously stored on the capacitor. Conduction through the resistor may thus be in either direction. The readout process is not entirely destructive, as the capacitor may not discharge too rapidly, and is only being discharged for a fraction of a microsecond; but regeneration of the information is normally done after writing anyway, simply by applying to the readout terminal whatever voltage was detected.

Typical operating parameters would be 8 volts across the two diodes in series, which produces a back bias of 4 volts on each of them if they are well matched. +2 volts on the capacitor then signifies a 1 state, and -2 volts in the capacitor signifies a 0 state. The two diodes are referred to as "squeezing" diodes, and the process of connecting them both to ground is referred to as "squeezing". Since the selection of a word in the dicap memory is performed by "squeezing" all the diode pairs corresponding to that word, dicap memories are inherently word-oriented. The readout terminal is also connected to a sense amplifier, of which there is one per bit of word length in a word-oriented memory.

This entire configuration would be quite suitable for fabrication using integrated circuit techniques. Quite possibly it could also be made to work well with more exotic comtemporary diode types such as "hot carrier" (also known as "Schottky-effect or "metal-semiconductor junction") diodes, tunnel diodes, or backward diodes, all of which have outstanding switching speed plus a much higher degree of radiation resistance than ordinary silicon diodes.

The diode-capacitor approach to information storage lends itself well to a bit-slice associative memory organization, and somewhat less well to a local logic organization. It might be possible to sum the current from two adjacent dicap bits in a single resistor; if the summing could be done with sufficient accuracy, local-logic operation according to signal cancellation techniques (Honeywell's "Scancell" examplifies this idea in the plated-wire case) would be a possibility. However, the precision requirements in this situation might be sufficiently so severe that the principal advantage of the dicap memory, its extreme speed, would have to be comprised. Thus, the most likely slot for a dicap memory is that of a small, extremely fast, bit-slice associative memory made with radiation resistance semiconductors. Such a memory would probably be excellent for use in the "lookaside" or "lookbehind" portion of a large-scale general-purpose aerospace computer, in addition to use with a much larger associative memory as was previously mentioned.

A technical reference on dicap memories is a short article by Arthur W. Holt, on pages 12-116 through 12-126 of the Computer Handbook, Edited by H. D. Huskey and G. A. Korn, McGraw-Hill Book Company, New York, 1962.

2.5 Ferroelectric and Ferrielectric Elements

A ferroelectric material is a material capable of being electrically polarized and of retaining this electrical polarization when the applied field is removed, just as a ferromagnetic material is capable of being magnetically polarized and of retaining its polarization when the applied field is removed. The basic physical explanation of ferroelectricity is quite different than that of ferromagnetism; ferroelectricity involves slight ion displacement in solids whose lattices have asymmetric unit cells with alternate positive and negative ions, whereas ferromagnetism involves lining up the spins of outer-shell electrons in certain substances where the effects of these electrons reinforce instead of canceling. Both phenomena are "cooperative" in the sense that, to achieve a permanent polarization which can be sensed microscopically, all of the small domains of polarization which naturally occur in the material must be brought to the same direction of polarization, so that they reinforce instead of opposing one another.

The normal "depolarized" state of either a ferroelectric or ferromagnetic is not one of neutral polarization at each lattice point; rather, there is complete polarization in each tiny domain, but the domains are randomly oriented and their effects do not reinforce and are not detectibal macroscopically. These domains are large enough to be visible using an optical microscope, if they are properly illuminated with polarized light; but they are far too small to be visible to the naked eye. Both types of materials may exhibit hysteresis, which is another way of saying that the internal polarization does not return to zero when the applied field does. In fact, the best materials of both types exhibit very square hysteresis loops.

For basic theoretical reasons from solid-state physics, all ferroelectric crystals are piezoelectric, although the converse statement is not true by any means. Thus, although there is also such a thing as a "zero-magnetostrictive" ferromagnetic material, there is no such thing as a "zero-electrostrictive" ferroelectric material. The term "piezoelectric material" is used more commonly than "electrostrictive material," but they are synonomous; however the term "magnetostrictive material" is used more commonly than its synonym "piezomagnetic material", so that usage in the two cases seems inconsistent. However, it really is consistent; the "piezo-" terms refer to a physical system in which pressure is the input and electric or magnetic flux change is the output whereas the "strictive" terms refer to the converse situation in which the converse physical effects appear. Any material exhibits the "piezo" effect if and only if it also exhibits the converse "strictive" effect.

Beside ferromagnetic materials, there are two other classes of magnetic materials exhibiting cooperative effects; these are respectively called antiferromagnetic and ferrimagnetic. In both of these cases, the situation is basically more complicated; the material behaves as if it is composed of two sublattices (or sometimes more than two), where each of these sublattices is magnetically polarizable, but in opposite directions such that their effects tend to cancel. If the oppositely directed polarizations do exactly cancel, the material is then said to be "antiferromagnetic," although this is not the same as being nonmagnetic because hysteresis effects are exhibited when the material is subject to a very intense field or

heated almost to its "Curie temperature." On the otherhand, if one of the sub-lattices is considerably stronger in terms of cooperative magnetic effect than the others, the material is said to be "ferrimagnetic;" such materials have hysteresis loops very much like these of ferromagnetic materials. To a first approximation, it is quite correct to say that a ferrimagnetic material is a magnetic material which is an insulator, although the physical situation is obviously by no means that simple.

The situation in the case of electrically polarizable materials is quite similar to that for magnetically polarizable materials, except that none of the former are conductors & all are insulators or dielectrics, and differ from other such materials in that they exhibit hysteresis. There are "antiferroelectric" materials, just as there are antiferromagnetic materials. There is also a class of materials which Professor C. F. Pulvari of Catholic University has described as "ferrielectric," although some workers in this field do not agree that he has conclusively demonstrated that these materials are not simply ferroelectric materials. The advantage of ferrielectric materials over ferroelectric materials, for computer applications, should in principle be that they exhibit a well defined "threshold field" - that is, there should be a certain finite electric field at which the material just begins to switch, and below which it will not switch at all. The ferrimagnetic ceramics, , called "ferrites," which are commonly used in magnetic computer memories have good threshold field properties. The underlying physical situation is considerably less clear in the case of ferromagnetic materials such as permalloy which are analogous to ferroelectric materials. In computer memory work these materials are usually presumed to have what is in effect a threshold field, but in practice such phenonema as "walking down" and "creep" are serious problems; these phenonema both have to do with the gradual degradation of stored information when individual storage elements are subjected to a relatively small field, less than would be required to write new information. "Creep" is, in particular, a serious problem in ferroelectric memories also.

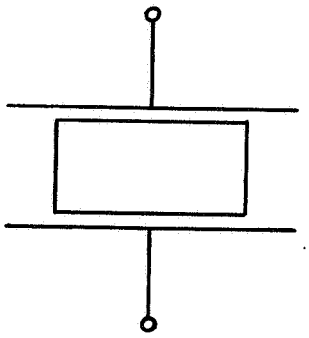
2.5.1 Word-Oriented

The simplest way to make a ferroelectric memory is to make discrete ferroelectric capacitors, and connect them up with some sort of selection electronics. The electronics may not need to be as elaborate as those used in a dicap memory because, unlike a linear capacitor, a ferroelectric capacitor will not discharge spontaneously if a circuit is closed connecting its terminals. Thus a drive line is used for all of the ferroelectric capacitors in a word, and then a second line is used for both driving and reading from all of the capacitors corresponding to one bit position in the memory. The use of a diode or two per capacitor may be necessary, if the threshold field properties of the ferroelectric material are relatively poor.

What is probably a much better type of ferroelectric memory can be based on the "transpolarizer" as developed by Pulvari. A description of the operating principles of the transpolarizer is given in an article by Pulvari, "The Transpolarizer: An Electrostatically Controlled Circuit Impedance with Stored Setting," C. F. Pulvari, Proceedings of the IRE, June 1959, pp. 1117-1123.

Readout from a transpolarizer storage element is in a sense pseudo-non-destructive (PNDRO) rather than fully NDRO, in that a PNDRO process partially destroys the remanent polarization pattern in a selected element,

but does not destroy the information stored in that element. It is possible after a PNDRO operation to simultaneously "regenerate" the information just read, or more precisely to "restore" the pre-readout polarization pattern everywhere in the memory, without affecting the state of any storage elements which were not read out. A magnetic storage element which operates in a rather directly analogous PNDRO mode is the "transfluxor," (see subsection 2.2.2), which has by now been used by many organizations in computer memories usually describes as "NDRO."

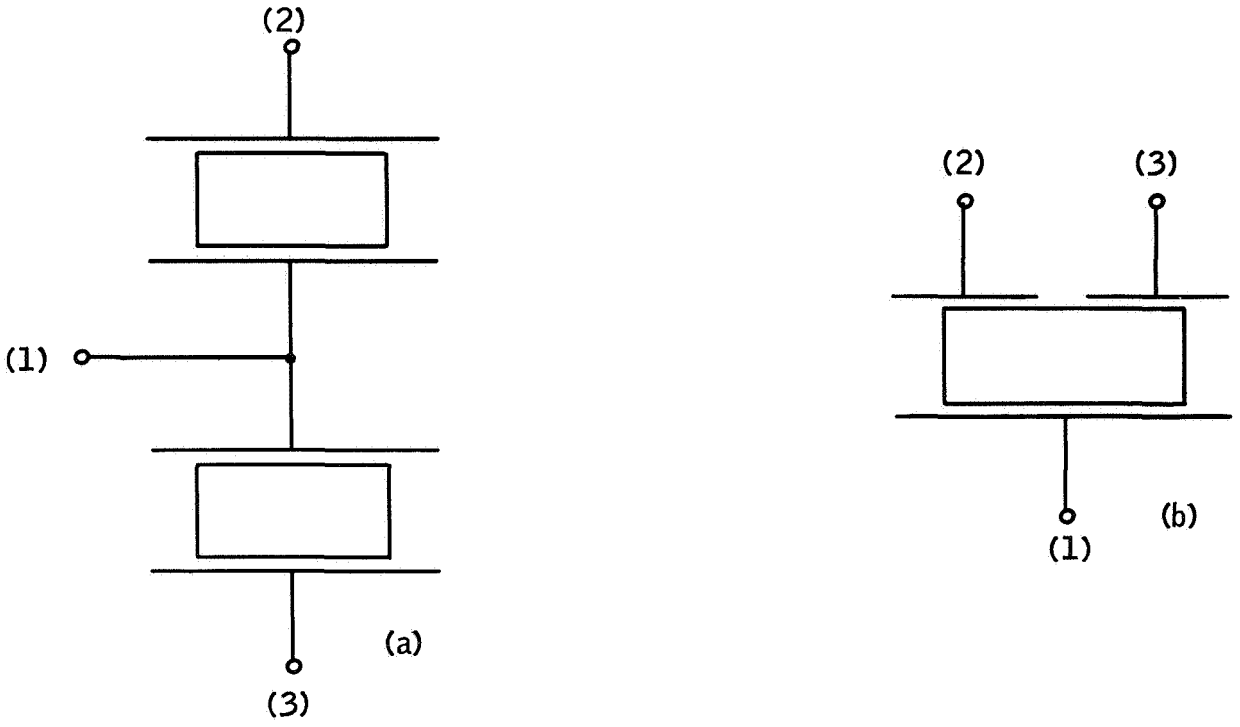


Symbol for a Ferroelectric Capacitor
Figure B1

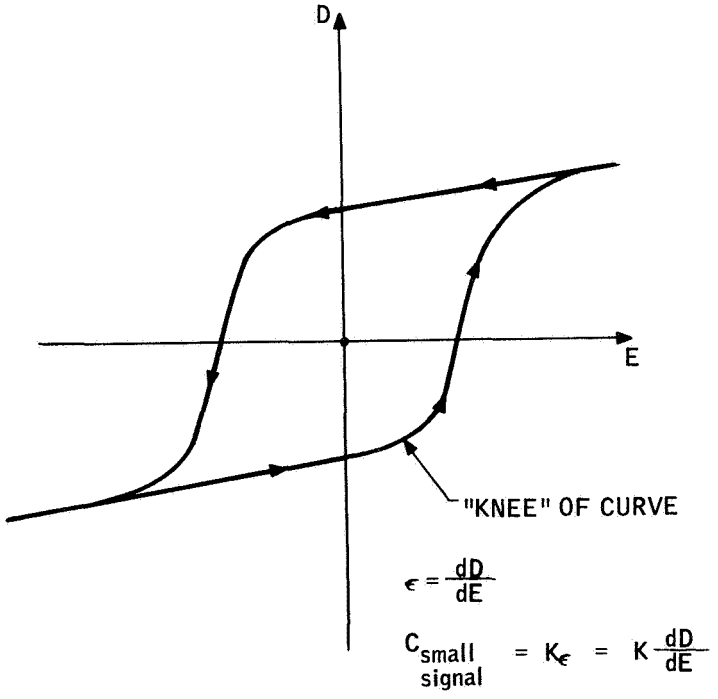
The usual symbol for a "ferroelectric capacitor" is shown in Figure B1; such a capacitor comprises a thin layer of ferroelectric material between two conductive plates. Such capacitors can be with the ferroelectric layer as thin as a quarter of a mil (or thinner), by use of radio-frequency ion-sputtering techniques. If thicker ferroelectric layers are desired, discrete ceramic chips can be ground down to a thickness of a few mills.

If now two ferroelectric capacitors are in a sense connected "back to back," the configuration of Figure B2(a) results. Figur B2(a) shows one symbolism for the transpolarizer; the actual geometry of the transpolarizer as made by Pulvari and his co-workers is closer to that suggested by the alternative symbolism of Figure B2(b).The transpolarizer in its simplest form is is a three-terminal device. For the purpose of storing new information in the transpolarizer, or of "restoring" information just read out, the center terminal (1) must be used; but the PNDRO process makes use only of terminals (2) and (3). Pulvari's standard device apparently consists of a round chip of ferrielectric material, with two conducting plates attached to each side would normally be connected togehter.

The basic hysteresis properties of the ferroelectric material are indicated in Figure B3. The "D-E" loop as shown is directly analogous to the "B-H" loop customarily quoted for magnetic material specimens. E is the electric field externally applied to the specimen, and D is the "electric displacement" or "total electric intensity" resulting within the specimen;

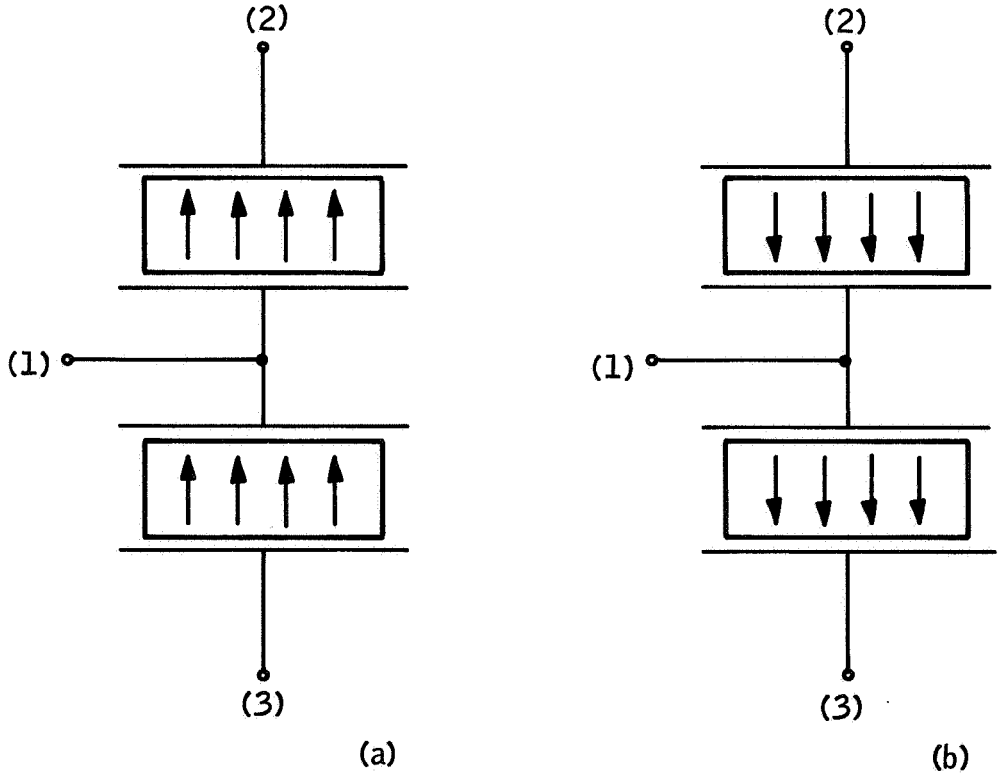


Symbols for a Transpolarizer Element
Figure B2



Ferroelectric Hysteresis Loop
Figure B3

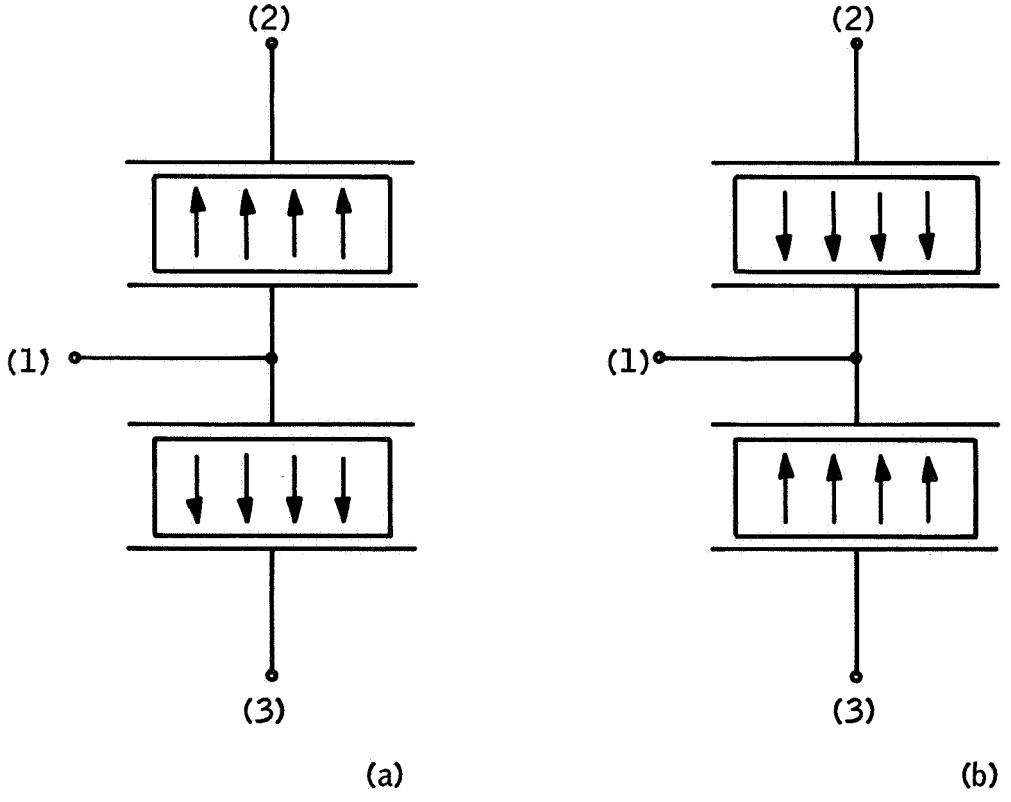
$D = P + \epsilon_0 E$ where P is the "internal Polarization" and may be several orders of magnitude greater than $\epsilon_0 E$. Hence the "P-E" loop usually quoted for ferroelectric materials is directly analogous not to a B-H loop, but to an M-H loop where $B = M + \mu_0 H$ and M is the "induction" or induced magnetic polarization. However, the difference between a D-E loop and a P-E loop for the same material may be indistinguishable to the eye on a graph drawn to ordinary scale. The slope $\epsilon = \frac{dD}{DE}$ of a line on a D-E plot is directly proportional to the small-signal capacitance C seen by an external circuit, where the constant of proportionality (here labeled K) depends on the system of units employed. The quantity ϵ is usually referred to as the "dielectric constant," although it is obviously not a constant for dielectrics which exhibit hysteresis (i.e., for ferroelectrics); "dielectric permeability" or "dielectric permittivity" are better names for E. ϵ_0 is "the dielectric permeability of free space, just as μ_0 is the magnetic permeability of free space.



Transpolarizers in Unblocked States
Figure B4

If both capacitors of the transpolarizer are polarized in the same direction, the transpolarizer is said to be "unblocked;" the two possible unblocked states are indicated in Figure B4. If, on the other hand, the two capacitors are polarized in opposing directions, the transpolarizer is said to be "blocked;" the two possible blocked states are shown in Figure B5. This terminology is identical to that in use regarding transfluxor states.

The meaning of the terms "unblocked" and "blocked" becomes quite clear when one considers what happens when a voltage is applied between terminals (2) and (3). Assuming that initially both capacitors are polarized in the same direction, and the transpolarizer is thus "unblocked," their capacitance then simply combine according to the "parallel resistors rule," and the net circuit element seen by the external drive circuit is simply one ferroelectric capacitor with a smaller effective capacitance than either of the two actual capacitors. If the latter are well matched, their apparent total series capacitance is half of the capacitance of either separately. If the applied voltage is of a polarity which would tend to reverse the polarization of each of the two series capacitors, switching will begin to take place. If, on the other hand, the signal is of the opposite polarity and thus tends to drive both capacitors further into saturation than they are already, the effect of the signal will be merely to superimpose an additional electric field on the remanent polarization; and this additional field will die out as soon as the externally supplied field is relaxed, without any permanent effect on the state of either ferroelectric capacitor.



Transpolarizers in Blocked States
Figure B5

Now consider instead the situation when the transpolarizer is in one of the "blocked" states in Figure B5. As the external voltage begins to build up between terminals (2) and (3), it is equally distributed at first between the two ferroelectric capacitors; thus it tends to drive one of them further into saturation than it already is, and tends to switch the other one. The latter capacitor, however, will eventually reach the "knee of the curve of Figure B3; when that happens, its capacitance (which is proportional to $\frac{dD}{dE}$) rapidly increases, so that the two capacitors suddenly have a grossly unequal capacitance. From this point on, most of the applied voltage will be felt across the first capacitor, since it effectively has a lower capacitance; and very little of the applied voltage will be felt across the second capacitor, which is the one capable of being switched in the direction of the applied voltage. The only effect of this voltage on the first capacitor is to drive it further into saturation. Thus until a much larger field is supplied, so that the second capacitor can be switched by the small fraction of the applied voltage which it receives, no further switching action takes place. The transpolarizer in this state is effectively blocked from switching, hence the use of the term "blocked" for the state. As is then quite natural, the other possible situation is referred to as the "unblocked" state.

PNDRO operation of an array of transpolarizer elements may thus be implemented as follows: A "zero" (0) is understood to be represented by a blocked state, and a "one" (1) by an unblocked state. The name "one-prime" (1') will now be understood to mean the distrubed unblocked state, which characterizes unblocked elements to which a reading voltage has been applied. This terminology is chosen here to be consistent with the usage in papers on transfluxor memories.

An attempt to read from a blocked element produces very little net current through the driver supplying the voltage, since the blocked transpolarizer behaves essentially like a relatively small linear capacitor. An attempt to read from an unblocked element will, however, produce an appreciable net current through the driver as some degree of partial switching takes place. This current signal, or its absence, is then detected by a special sense amplifier designed to read this type of signal, and is accordingly interpreted as 1 or 0. After the transpolarizer has been read from, it is "restored" by applying a voltage pulse similar to the read pulse, but the opposite polarity. The restore pulse is applied across terminals (2) and (3), as was the case with the read pulse.

The advantage of this PNDRO operation, over conventional DRO operation, is that a large number of words may be read out from a PNDRO memory without restoring them individually after reading. When the PNDRO reading process has been completed, the words read out can be all restored simultaneously, without there being any need to explicitly reobtain the information previously read from them and use this information to control the "bit drivers" or "inhibit drivers" as in DRO memory operation. The reason that this simultaneous restoration operation is possible is that a "restore" signal will not affect any elements except those which are in the 1' state; those in the 0 state are "blocked" with respect to an applied voltage of either polarity because the transpolarizer effect is symmetrical, and those in the 1 state can merely be driven further into the 1 state by the application of ther restore signal. Hence, only elements in the 1' state can be affected by the restore signal; and these will be partially switched, sufficiently to return them to saturation in the 1 state direction.

The transpolarizer is in principle an eminently suitable device for an associative memory. The transpolarizer would probably be fairly well suited, just as the transfluxor is, to either a bit-slice organization with external logic or to a local logic organization.

2.5.2 Coincident Voltage

The classical ferroelectric memory configuration is a coincident voltage scheme which requires a material with a very square loop and a very high resistance to creep. Suitable materials however did not exist at the time most of the work was done on this type of storage. Now that they do exist, the state of the storage art in magnetics has progressed to the point where ferroelectric stroage of this type is probably of no longer much interest. A good account of the principles of ferroelectric storage is to be found in Digital Computer Design, Edward L. Braun, Academic Press, N.Y., 1963, pp. 253-256.

The basic idea is that parallel electrode strips are deposited across a thin wafer of ferroelectric material in one direction on top and in the other direction on the bottom. If E volts are required to switch the material, and the hysteresis loop of the ferroelectric material is sufficiently square, it will be insensitive (except for creep) to 1/2E volts and thus if + 1/2E is applied to a given upper electrode, and - 1/2E to a given lower electrode, the only bit position in the wafer which will have an applied voltage of E is the one at the intersection of the two energized electrodes.

Unfortunately, there will be some spreading of these fields within the material, and therefore some interference between bits. Various schemes can be employed to try to reduce this effect; apparently the standard one is to also put a voltage of 1/2E on all the other lower electrodes parallel to the one that has the voltage of -1/2E, which means that there is no net voltage between the selected electrode on top and the unselected electrodes on the bottom. Obviously this modification is not a complete cure to the problem, because there are still unselected upper electrodes which are at ground, and there is a field of 1/2E between each of these and each unselected lower electrode.

The outstanding potential advantage of this type of memory, which originally attracted the attention of researchers in this computer field as early as 1952, was its very small size as compared to other early computer memories. Now, however, core memories have become very much smaller, and other types of compact batch-fabricated memory systems have emerged which appear more promising than coincident-voltage ferroelectric memories.

2.6 Electro-Optic Media

An electro-optic memory typically consists of a continuous plane of optically sensitive material, together with some light source which can be switched to illuminate and thus to "select" any position on the plane being used for information storage.

Optical selection techniques are quite varied. There may be strips of light-source material physically adjacent to the plane itself, or in another plane at some distance from the storage plane, with projection optics in between. Alternatively, there may be a fixed light source with an electro-mechanical scanning mechanism, and in a few years, there may also be satisfactory solid-state light-beam scanning devices. Another possibility is a geometrically extended continuous light source, one small portion of which is caused to turn on at a time by the application of some sort of electrical or electro-mechanical delay signal which propagates along a medium. The optical scanning mechanism could even consist of a cathode ray tube with a electroluminescent face, or of some other type of bright display divice.

The optically sensitive materials that are used should undergo a marked change in some physical property upon stimulation by light. This property should be one which in turn results in electrically detectable changes or, in some cases, optically detectable changes in the information storage material, which may or may not be the same as the optically sensitive material. The most commonly used sensitive materials are photoconductors whose conductivity may increase by up to eight decimal orders of magnitude (cadmium selenide) upon sufficient illumination. There are other materials in which the light produces irreversible chemical changes where it is applied, resulting in the projected image becoming "fixed". There are still others where applied light of one frequency can produce "color" (not necessarily in the visible spectrum) changes, which can be detected by illuminating the storage material with light of another frequency.

The number of possible electro-optic memory schemes is essentially unlimited. Some of them feature both optical write-in and optical readout; probably a larger proportion of them feature optical write-in and electronic readout. The third logical possibility, electronic write-in and optical readout, is also represented in the list of devices in this section; but such devices are usually described as "displays," even though they may also be of value as "memories."

Some general comments can be made about the use of electro-optic media as associative memories. First of all almost any electrical-optical medium can be configured in a way analogous to a conventional word-oriented or "linear select" organization for magnetic core or thin-film memories. Whereverrthis is possible, a bit-slice organization is also possible, and it will be useful for associative memory implementation if the readout process is non-destructive and is sufficiently rapid. Secondly, if the signals produced by this type of NDRO operation are sufficiently uniform when measured over every information storage element of the storage medium, the possibility also exists of performing "local logic" by means of signal cancellation techniques similar to those proposed for Honeywell's plated-wire SCANCELL element. However there are other operating modes of an electro-optic memory which do not have an equally direct analogy in other types of memory and which would be worth serious consideration wherever the job to be done consists primarily of equality searches and/or proximity searches, and the speed requirements

are notably high. In one such mode, the electro-optic associative memory becomes, in effect, an "area correlator" which directly correlates one "picture" (a centrally supplied search word in the form of a bright display) against the local "pictures" supplied by each sub-area of memory. In general, the way this may be done is that the polarities and information representation conventions are set up in such a way that a non-zero output signal will be produced only upon a "mismatch" or "anti-coincidence," that is (a 1 state in the search word opposite a 0 state in the memory word or vice versa), and there will be no detectible signal out in the event of a "match" or "coincidence" (a 1 or 0 state in the search word opposite an equal state in the memory word). Thus, a search word might for instance be organized in a square array of 8x8 bits, to be matched against every 8x8 bit area on the surface of an electro-optic memory plane; a complete match for one 8x8 area would then produce no output signal from that area at all, whereas a mismatch would produce an output voltage or current signal roughly proportional to the number of mismatched bits. Similar phenomena may be made to occur, of course, in memories which are not electro-optic; but the difference here is that exposure of the electro-optic medium to the search word can be very conveniently performed as a completely parallel operation over all bits of all words in the memory. Moreover, some electro-optic storage media have the capability of storing, not just 1 and 0 states at a given point, but a multiplicity of "gray-scale" levels; an electro-optic memory making use of gray-scale recording would fall in between what is normally called an "associative memory" and what is normally called an "area correlator". A second possible operating mode may be defined in which all of the words in memory, rather than just the search word, are stored as bright displays; that is, the information storage medium itself produces either a light output for a 1 state and no light output for a 0 state or else the converse arrangement. In this case, the optically sensitive plate would not be used as an information storage device itself, but as a continuous logic device somewhat analogous to the "detection plane" in certain more conventional associative memories. For instance, the search word may be stored as a bright display at one optical frequency, and a word in memory may be stored as a bright display at another frequency; and by the use of an optically sensitive plate coated with two photoconductors having non-overlapping spectral sensitivity characteristics (such as cadmium sulfide and cadmium telluride), the stroage medium behind the photoconductor layers may be made to change state if and only if light of both frequencies is received at a given spot. To perform the usual form of equality or proximity search operation it would then be necessary to store a bit of information in both "true" and "conplemented" form, both in the search word and also in the main memory, so that matching could be implemented as the usual sort of "exclusive-or" "anticoincidence" logic functions.

Various media will now be discussed briefly, keeping these different forms of associative memory organization in mind.

2.6.1 Ferrotron

"Ferrotron" is the name given to an electro-optic image plate having a photoconductive layer as the selection element and a ferroelectric layer as the information storage element. A fairly complete introduction to ferrotron memories is given in Honeywell Systems and Research Division Technical Report R-RD 6373, An Electro-Optic Computer Peripheral Mass Memory With a Removable Data Cartridge, C. W. Hastings, 18 August 1965. The basic description portion comprises pages 1-7; tne balance of the

report deals with possible optical selection schemes which are based on electronically selected light sources, projection optics, and the use of two photoconductive materials having non-overlapping spectral response characteristics.

A ferrotron memory in simplest form consists of a continuous sheet of ferroelectric material on a conducting substrate, overlaid with a coating of photoconductive material, which is in turn topped off by a layer of transparent conductive material such as NESA (SnO_2). Assume now that the entire ferrotron plate is dark except that one small spot is illuminated, and that a voltage is applied between the upper and lower conducting electrode layers. This voltage will essentially be applied entirely across the photoconductor except right at the illuminated spot, since the resistivity of good photoconductive materials is extremely high in the absence of light. At the illuminated spot, however, the voltage is felt almost entirely across the ferroelectric. A voltage of suitably chosen magnitude can be made, in this way, to polarize a small spot of ferroelectric material into saturation in either the "up" or the "down" direction. In this manner a "charge pattern" can be stored on the ferrotron plate by means of a beam of light.

It should be emphasized, however, that this "charge pattern" is really a hysteresis-effect remanent polarization pattern, and that it will not "leak off" as if the ferroelectric plate were an ordinary capacitor. The observable polarization is not actually due to the presence of any excess domains in the material, in similar fashion to the lining up of small homogeneously polarized magnetic domains in ferromagnetic and ferrimagnetic materials such as permalloy tape-wound cores (subsection 2.3.3) and toroidal ferrite memory cores (subsection 9.1.1).

Once the polarization pattern has been stored, it may be read destructively by scanning with a beam of light and maintaining a DC voltage between the plates in the direction suitable for writing a "zero". Since the amount of current flowing through the voltage source will be negligible (for square-loop materials) when the light beam is scanning a "zero" and appreciable when it is scanning a "one", a current-detecting sense amplifier in series with the DC voltage source can be used for readout.

The ferrotron is an example of an electro-optic storage medium in which the information would normally be written optically, and read back out electronically. A ferrotron thus is a "video transducer". Ferroelectric layers can, however, also be used to control electro luminescent layers for display purposes; this "ELF" (electroluminescent-ferroelectric) arrangement has been extensively investigated by Westinghouse. Thus the possibility at least exists of a ferrotron memory in which both write-in and readout are optical. (One could conversely consider the "ELF" device class as an electronic write-in, optical readout memory.)

The Honeywell Systems and Research Division is currently investigating the detailed electrical behavior of ferrotron devices and of ferroelectric materials, and studying approaches to fabricating ferrotron devices from some of Honeywell's ceramic ferroelectric materials.

One early result of this work has been a scheme for combining the ferrotron and transpolarizer (see subsection 2. 5. 1) principles by using a ferrotron plate having narrow parallel strip electrodes in one direction on the underside.

Figure 6 shows, in cross section, one possible configuration of a plane of ferrotron transpolarizer elements. Two adjacent electrodes define the same block of data, since two ferroelectric capacitor elements per bit of storage are now required instead of one. The "transparent conductive layer" of Figure B6 corresponds to the terminal (1) of Figures B2, B4 and B5; it plays an essential role during the writing of new information, and is a "shunt" during reading. Reading is then accomplished by applying a voltage between the two metallic electrodes which correspond to terminals (2) and (3) of Figures B2, B4 and B5, and illuminating the photoconductor. Obviously, the sneak conductance parallel to the plane through the ferroelectric material must be very low, relative to the conductance vertically through the ferroelectric and photoconductor layers. * A transpolarizer element is thus selected for reading by the coincidence of a light beam illuminating the photoconductor material above two electrodes, and a driver applying a voltage between the two electrodes. When a block of data has been read by sequentially scanning the light beam along the electrodes, restoration of the information just read can be performed very easily by reversing the direction of the applied voltage (which will require a driver circuit of somewhat unusual capabilities) and then briefly illuminating the entire ferrotron plate. There will be no effect on any of the unselected elements of the plane, because their corresponding drivers are not activated; and there will be no effect on selected elements in the 0 or 1 state, for reasons previously discussed. Hence only those elements in the data block just read and left in the 1' state will be affected.

2. 6. 2 Silver Plate Sandwich

The silverplate sandwich memory is a development of Honeywell Radiation Center, Brighton, Massachusetts. Related work has been done at other companies, notably at Philco. In Honeywell's silverplate memory, the information storage plane consists of a sandwich similar in structure, to a ferrotron plane except that in place of a ferroelectric material there is a thin layer of spongy black material (such as black blotting paper) soaked with a silver-plating solution (such as silver cyanide or silver iodide). In between this spogy layer and the photoconductor, there is a thin layer of silver. Both the front and the back electrodes of this type of memory plane must be transparent. Recording is performed much as in the ferrotron; voltage is applied in the direction which would tend to plate silver through the spongy layer between the front and back electrodes and the plate is then exposed to a projected image of a bright display of the information to be recorded. Plating through will occur wherever the photoconductor in front is exposed to enough light that it becomes conductive, and it will not occur anywhere else. Wherever silver plating occurs, the back of the plate will henceforth (until erased by a writing process with oppositely directed voltage) appear silvery and reflective; wherever plating does not occur, it will remain optically absorbent or black.

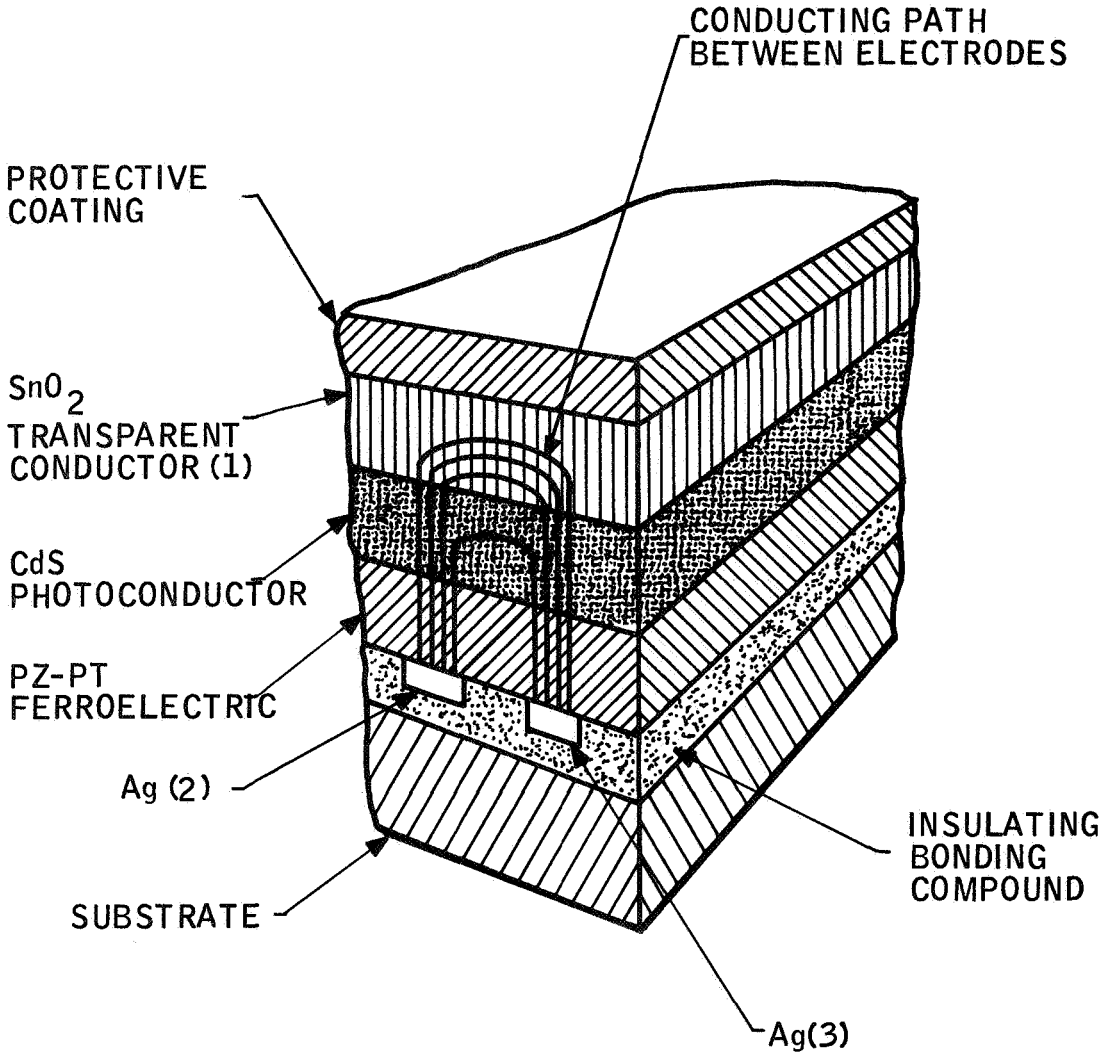


Figure B6. Ferrotron Transpolarizer Plate - Sandwich Structure

If the back of the plate is then illuminated by a bright light source, the reflection will be relatively intense wherever the plating through action has occurred and negligible elsewhere. Consequently, the silver plate memory is well suited to both optical write-in and optical readout. It amounts to a kind of re-useable digital photographic film; one can optically read out an image of an entire two-dimensional area of the silver plate, and project this image onto some receiving buffer storage element such as a ferrotron plate or VIDICON or ORTHICON tube. Considering the number of bits that may be stored in a small two-dimensional area and putting the rate of information transferred by this direct optical image method in terms of characters per second, the resulting number will greatly exceed normally encountered electronic information transfer speeds. It might, for instance, possibly require less than 10 microseconds to transfer a block of 4,096 8-bit characters or $2^{15} = 32,768$ bits, which would occupy an area on the ferrotron or silverplate memory plane of a fraction of a square inch. There is reason to expect information packing densities of 500 to 1000 bits per inch in both directions with the silverplate memory, just as with the ferrotron memory.

The system operation of a silverplate mass memory might proceed as follows: Duplicate images of a block of information to be stored would be projected on many different portions of the silverplate simutaneously by means of "kaleidoscope" optics similar to those used to project a mask pattern onto many sub-areas of a silicon chip simultaneously during the manufacture of semiconductor devices. There would be independent electronic selection for each of these blocks, such that only one of them would have a voltage between the electrodes capable of causing plating to take place. The back side of the silverplate would face another set of kalaidoscope optics; one electronically selectable light source would illuminate each entire block of information, and each light source would be shielded from all other portions of the plate. The selection of one of these light sources would thus porduce an image of the silver deposition pattern by reflected light, which could be transferred to a video sensor such as a VIDICON or ORTHICON tube, or a ferrotron plate. Conversion of the information to the usual sequential electronic logic form would not be performed within the silverplate mass memory itself.

For the purpose of writing into the memory, a bright optical image of the information to be written would be required. This image could be obtained by using a cathode ray tube, or else a gas-cell display panel similar to that recently announced by Lear-Sigler Corporation (subsection 2.6.6) which in effect has 20 tiny gas discharge tubes to the inch in both directions, directions, all independently controllable by a coincident-voltage technique. Also, it is now possible to purchase matrices of light-emitting diodes in configuration of up to 192 diodes, and such a diode matrix could be used iteratively to record portions of the block sequentially.

2.6.3 Photoemulsion

The storage medium used for the mass memory of a file computer under development by Itek Corporation, Lexington, Massachusetts is a ten-inch diameter plastic disk, coated with a high-resolution emulsion. This

emulsion has most of the properties of an ordinary photographic emulsion; but it also has the interesting additional property that even after it has once been exposed, developed and fixed, it can be essentially "unfixed," re-exposed, and redeveloped and the information it will then contain will be the logical "or" of the two exposures. The number of exposures is in fact not necessarily limited to two. Thus, additional recording and up-dating of files is possible, although erasure of information already recorded is not. These disks are stored in horizontal stacks; they are selected electromechanically by techniques resembling those used in a conventional audio jukebox, placed on turntables, and read using an optical-to-electronic transducing system consisting of a cathode-ray tube, a lens, and a photomultiplier tube. Writing is performed by a digitally controlled 27-milliwatt laser.

The Itek file computer configuration would not be of interest for associative memory applications. However, quite possibly it represents the largest-scale device in existence at this time; Itek claims that a complete system with a full complement of equipment will have a storage capacity of more than a trillion bits, which is roughly a thousand times the capacity of the largest disk file memories now for sale. However, the ability of Itek's memory medium to record new information additively could make it of interest, at least if implemented with some non-mechanical scanning principle, in situations where a typical read-only mechanically-alterable memory would not be of interest.

In any case, Itek's memory differs from those discussed in all other subsections and sections of this document in that the medium itself is transparent, and the reading and readout is by transmitted light rather than by reflected light or by light controlled by the information itself. A second interesting transmitted-light bulk storage system, known as the "videofile" system is under development at Ampex Corporation, Redwood City, California.

This system is described in the Record of the Western Electronic Convention, August 24-27, 1965, in booklet section 18, entitled Advanced Techniques and Memory Designs. The recording medium used here is photographic film having a mylar base, a silver halid emulsion, and a thin conductive layer above or below the emulsion to facilitate charge dissipation, which would otherwise be a problem with the electro-beam recording technique. Recording in this instance is permanent, and is performed using electron beam techniques.

There have been other photo-emulsion recording media, and also other media having similar characteristics such as General Electric's thermo plastic cards, developed for information retrieval applications. All media of this type would be most useful in read-only memories; the Itek medium is the only one which could be used in any other way. The recording beam scanning and record selection techniques used in these bulk storage systems are generally at least partly electro-mechanical, and would be quite unsuitable for use in an aerospace associative memory. Thus the development of associative memory based on these media hinges partly on the development of precise techniques for deflecting light beams and electron beams.

One intriguing alternate possibility is the use of a fiber of material with low acoustic loss (one for each memory plane), coated successively with a non-linear resistance layer, an electroluminescent layer, and an outer conductive layer. In this sort of arrangement, an acoustic pulse may be applied to one end of the fiber by a piezoelectric transducer. As the pulse propogates down the fiber, it will cause the electroluminescent material to light up as it passes each area, if the voltage between the electrodes is maintained at a suitable value. This type of scanning device could be coupled with a cylindrical diverging lens to spread the light produced to span an entire memory plane. If the fiber lines in the direction corresponding to the "bit direction" in the memory medium, bit slices will be illuminated one after another.

2.6.4 Photochromic Media

Some materials slowly change color upon irradiation with one wavelength of light, and may be scanned with a different wavelength. Notable among these is the mineral hackmanite ($6NaA1SiO_4 . 2NaC1$) which has a natural pink color which is lost upon irradiation with sunlight. Subsequent exposure to ultraviolet radiation produces a deep violet color, which may be similarly bleached by sunlight. Material in this class are also termed "tenebrescent". Certain other metallic salts exhibit a "color center change" which can be used to store optical images with extremely great resolution. Work in this area has been, and is now, in progress under Air Force sponsorship at various laboratories.

There are also liquid dyes which exhibit phototropism, or the change of color, upon irradiation. These may be encapsulated in gelatin, using techniques developed by National Cash Register, Dayton, Ohio, to produce a usable material which can be made into a medium on a plate almost as though it were a powder.

The problem with all of these techniques is that both read-in and read-out must be optical. The development of read-out scanning techniques sensitive enough to pick up output signals corresponding to changes in color of the memory plane is probably a more difficult task than either obtaining suitable memory plane materials or developing methods of recording information on them. References on memory applications of hankmahite are: Investigation of Inorganic Phototropic Materials as a Bioptic Element Applicable in High Density Storage Computer Memories, Polacoat, Inc., ASD TR 62-305, April 1962; and Research on Automatic Computer Electronics, Lockheed Missiles and Space Company, RTD-TDR 63-4173, pages A-153 through A-155; February 1964.

2.6.5 Phosphor-Photoconductor Latch

A "sandwich" of an electroluminescent material or "phosphor" such as zinc sulfide or cadium sulfide, and a photoconductive material which may be the same or a related compound, can be made to form a volatile bistable element, or "latch." This type of element has sometimes been called an "optron". Once current has been applied to the phosphor and it lights up, any resulting emitted light falling on the photoconductor may cause it to conduct sufficiently that the current controlled by the photoconductor may, in turn, be used to keep the phosphor producing light. After the optron element has been turned

on, it may be turned off by interrupting its current supply and turned on again by applying a suitable light pulse to the photoconductor.

The advantage of this kind of device for some associative memory applications would be that it produces light as an output; if the images of an optron storage plane and an optron display of a search word were projected on a "detection plane" consisting of some other kind of electro-optic medium, in particular a ferrotron, local-logic equality or proximity search operations could be implemented. The principle disadvantage of optrons as memory devices is their relatively slow operation, apparently down in the tens of kilocycles region.

A reference on optrons is Electroluminescence, H. K. Henisch, Pergamon Press, Inc., New York, pages 277-288; 1962.

2.6.6 Gas Discharge Cell

An unusual device developed by Lear-Siegler Corporation, Grand Rapids, Michigan as an "energy management" cockpit display for the X15 vehicle consists of a 4x4 inch glass sandwich panel less than a quarter of an inch thick. This panel has 20 tiny gas cells per inch in each direction, which are electronically selected and controlled by a coincident-voltage technique. The top plate of the sandwich has transparent anode strips which are oriented in one direction, and the bottom plate has cathodes which are opaque and reflective and are oriented in the orthogonal direction. The middle of the sandwich is a glass plate with square holes etched through it, each of which defines one gas call. A cell which has been off (non-conducting and unlit) may be turned on by a voltage of 285 to 295 volts between its anode and its cathode; the light output can then be maintained with only 195 to 200 volts anode-to-cathode. It takes about 10 microseconds to turn on a cell which is off, but it takes about 500 microseconds to turn off a cell which is on. Qualitatively, the electronic operation of each cell is quite similar to that of the silicon-controlled switch in this respect (subsection 1.1.4).

This display device might potentially be used as an electronic-input, optical-output digital memory. The write time of a previously cleared or "blank" memory plane is effectively only 10 microseconds although a random-access write time would have to be quoted as 500 microseconds, since cells which were on might have to be turned off. The light output produced is sufficient to make the display visible even indirect sunlight.

Electronic operation of the gas cell array as a memory might be achieved as follows: Unselected anodes would be maintained at approximately 30 volts, and unselected cathodes would be maintained at approximately270 volts. Thus, all unselected elements would have 240 or 270 volts across them, which would be sufficient to keep them if they had been turned on, but not sufficient to turn them on if they had been off. To select a cell to be turned on the appropriate anode would then be grounded, and the appropriate cathode potential would be raised to 300 volts, thus applying 300 volts across the cell at the intersection of the respective anode and cathode, which would be sufficient to turn it on. To select a cell to be turned off, its anode voltage would be increased to 65 volts and its cathode

voltage would be reduced to 235 volts, so that the voltage across it would drop to 170 volts long enough to extinguish conduction. All other cells would have 205 or 240 volts across them, which would maintain whatever state they were in. This operation is somewhat similar in principle to that of the coincident-voltage ferroelectric memory previously described in subsection 2.5.2 with the difference that readout is not performed by a technique related to that used for writing, and the "field-spreading" problem is of no consequence because of the electrical isolation of each cell from every other cell.

The engineers who developed the device at Lear-Siegler feel that information densities up to 50 cells per inch in each direction are ultimately practical. The complete memory is said to be very inexpensive; its only serious problem is volatility in the sense that the information is lost if power is lost.

A possible high speed equality - and proximity-search associative memory would use a gas cell display to store information, and use a ferrotron plate with two distinct photoconductor layers as a detection plane as already discussed at the beginning of section 2.6.

3.0 Read-Only Elements

The term read-only element here signifies "an information storage device which is considered to have a 1 state or 0 state pre-stored in it at the time it is fabricated and has no electronically controllable means of changing state. A "card changeable" element is a read-only element packaged in such a way as to be conveniently alterable by the computer system human operator.

Most existing read-only elements which might be considered for associative memories are discussed in reference 2. Ferroelectric read-only memories and serial-access fixed memories were not covered in this reference, nor were optical read-only memories; the latter have been covered briefly in subsections 2.6.3 and 2.6.4 of the present document. Except for optical read-only memories, the read-only element generally comprises some sort of coupling between a drive line and a sense line. This coupling may be by inductive effects, by capacitive effects, or through a non-linear element such as a diode. The inductive and capacitive cases may be further subdivided according to whether the coupling is simply across an air gap, or is enhanced by a small amount of hysteresis material (ferromagnetic or ferroelectric, respectively in the two cases). From this point on, the geometries and operating modes of read-only memories show a bewildering and divergent variety.

Read-only memories are usually subdivided into three major categories: non-alterable memories, mechanically-alterable memories (which nowadays are sometimes called "card-changeable" memories), and electrically-alterable memories (which are simply NDRO devices based on hysteresis elements). If one is specifically interested in using existing types of read-only memories for table-lookup applications, this classification appears to be a most useful one. On the otherhand, if the end in view is that of discussing read-only associative memories which may be available in ten years, the most useful classification is probably the more macroscopic one in terms of the physical basis for the information selection and reading phenonemon, since almost all other characteristics of read-only memories are apt to change considerably within the next ten years.

There are, of course, certain memories which appear to fall in between some two of the three archtypes "non-alterable," "mechanically alterable", and "electrically-alterable". For instance, the mechanical design of an electrically-alterable memory may imply that the circuits which enable writing in the memory are physically detached from the memory itself during system operation; after the memory has been loaded and the write circuits have been unplugged, the same memory becomes for system purposes "non-alterable." This approach is commonly taken in the design of memories for aerospace equipment, in which there is the added motivation of economizing on size, weight, and power consumption of the avionic memory by leaving the write circuits on the ground.

None of these memory archtypes is unequivocally superior for all table-lookup circuit applications. In general, it is much safer to use non-alterable memories where the information to be stored is mathematically defined, and there is little risk of having to redesign the entire memory circuit because of corrections or changes in this information. Once a program has been very thoroughly checked out and is therefore believed sound, it may be desirable for economic or reliability reasons to use a mechanically-alterable memory for program storage instead of an electrically-alterable memory, as long as the future changes in the stored information are not expected to be so frequent as to be unduly bothersome or expensive. Electrical alterability is by no means a desirable property per se, since it may imply the possibility of inadvertently destroying necessary information in the event of program derailment, when a misinterpreted instruction writes meaningless information back into the program memory and in doing so "clobbers" part of the program. It is obviously a safety factor, in an avionics system whose correct operation depends entirely on an aerospace digital computer, that there be no capability aboard for writing into that portion of the computer's program memory which contains the stored program and/or the stored microprogram. Moreover, it is probably correct to generalize that mechanically-alterable memories cost less per bit than electrically-alterable memories, even though there are many applications where nothing less than an electrically alterable memory will do.

When the application involves an avionic associative memory, there may still be circumstances where electrical alterability is not warranted, and where a mechanically-alterable memory could be made smaller and lighter, would consume less power, and would afford greater protection to vital mission data against accidental destruction or compromise due to program or circuit malfunctions during the mission, than an electrically-alterable memory. Typical mechanically-alterable memories are "card-changeable," which is to say that the removal of a card of some type and the insertion of a different card suffices to change the information content of a portion of the memory; hence, the memory still has the capability of being reloaded with different information conveniently for the next mission. Some suggestions of possible system roles for read-only associative memories are given on pages 2-32 through 2-34 of reference 1.

The last category of read-only memory to be considered, serial-access, is a new idea. Such a memory would consist of a "strain-wave memory" (see subsection 4.2.1), fabricated in such a way that the information context is stored permanently during fabrication as the presence or absence of hysteresis material at each point along a memory line.

3.1 Inductively Coupled

Inductively coupled fixed-information memories may use conventional word-oriented ("linear-select") organization, conventional coincident-current organization, or different organizations which cannot be readily identified as either of these. In order to be able to use an organization other than word-oriented, it is generally necessary to having coupling by a hysteresis element.

3.1.1 Ferromagnetic or Ferrimagnetic Coupled Inductive

This type of memory has reveived more research and development attention than any other type of read-only memory. The simplest form it can take is an array of drive wires crossed with an array of sense wires, with the presence of a magnetic hysteresis element at an intersection signifying a 1 state and the absence of such an element signifying a 0 state; such a memory is often called a "slug memory." A mechanically alterable slug memory was constructed by Ferranti, Manchester, England, for the ATLAS computer, which is a British STRETCH-class system of which there are by now several copies in existence. The standard ATLAS slug memory (called "the fixed store" in Ferranti's terminology) consists of 8192 48-bit words, accessible on a random-access basis at a parallel read repetition rate of 4 megacycles. The original design goal was 5 megacycles, but 4 megacycles is apparently attained in normal operation of the computer.

Basically, the ATLAS slug memory is a rather specialized 48-bit, word-oriented ferrite, random-access memory. The sense wires run in the "bit direction," and the drive wires run in the "word direction." When a word wire is driven, a much larger sense output is detected in those sense wires corresponding to bit positions in the driven word where there is a ferrite slug at the intersection of the word wire and the sense wire, than is detected in those sense wires which do not correspond to slugs in the selected word. (Some obvious variations of this scheme would permit "bipolar" sense output waveforms if they are desired.) The physical form of the ferrite slug is that of a small peg which can be inserted into a hole on a board. Ferranti's literature indicates that the pegged-up fixed-store program is to be considered as part of the hardware delivered by the manufacturer, and is not normally to be changed after an ATLAS has been installed. A reference is "A Digital Computer Store with Very Short Read Time," Proceedings of the Institution of Electrical Engineers (Great Britain), Volume 107B, pp. 567-572, November 1960.

A different memory configuration, which can still fairly be considered a "slug memory," was developed for similar system applications in the CIRRUS digital computer. This memory comprises 4096 40-bit words, accessible on a random-access basis at a read repetition rate of 667 kilocycles. It is apparently much lower in cost than the ATLAS memory. Alteration of the information in one of these words would require rethreading the drive wire for that word. CIRRUS is a scientific computer developed by the Electrical Engineering Department of the University of Adelide, Adelaide, South Australia. References concerning CIRRUS and its memory are: "An Economical Multiprogram Computer with Microprogram Control, "M. W. Allen, T. Pearcey, J. P. Penny, G. A. Rose, and J. G. Sanderson, IEEE Transactions on Electronic Computers, pp. 663-671, December 1963; "A Pre-wired Storage Unit," I. R. Butcher, IEEE Transactions on Electronic Computers, pp. 106-111, April 1964.

The "core rope" is probably the most widely used type of non-alterable magnetic memory. Considerable work on core ropes has been done by the Honeywell Electronic Data Processing Division and Raytheon in the Boston area, and by the Burroughs Electronic Components Division, Plain Field, New Jersey. An article surveying the core rope state-of-the-art was published last year, by two ex-Burroughs men ("Application of Rope Memory Devices," D. Clemson and P. Kuttner, Computer Design, pp. 12-22, August 1964.) Core rope memories are offered on a commercial basis by Burroughs. Core ropes have been used in several computers (H-290, H-400, etc.) built by the Honeywell Electronic Data Processing Division, and in the aerospace computer built by Raytheon for the Apollo program. The Honeywell Systems and Research Division recently constructed two relatively high-speed core ropes; each effectively comprises a small random-access memory which stores 238 words of 8 bits plus parity, and which has a readout memory cycle time of 1.33 microseconds.

These ropes use 50-mill tape-wound cores each having 20 turns of 1/8-mil permalloy tape; the wiring is laid out in a square array in order to optimize noise cancellation, instead of in the usual "rope" configuration.

Core rope memories have one clear-cut design flexibility advantage over any other type of random-access memory, although this advantage is not of consequence in the vast majority of applications: they allow complete convenience and efficiency in the use of a "non-compact address." The meaning here of "compact address" is an s-bit binary number which is used to address one of 2^s distinct cells in a memory; a "non-compact address: is, logically enough, an s-bit binary number used to address one of N cells in a memory, where N is smaller than 2^s. In general, N need not be a power of 2 and may sometimes be many binary orders of magnitide smaller than 2^s. The justification for the use of the word "compact" is that, from first principles, no reduction can be made in the number of bits in a compact address without the number of distinct representable address states becoming less than the number of cells to be addressed.

The reason that core rope memories are efficiently addressable by a non-compact address is that, as will be explained shortly, address decoding takes place "locally" in each core of the core rope - and thus only those particular combinations of the s address bits must be decoded which are explicitly desired to correspond to a core, and consequently to a cell. More over, the decoding is determined by the wiring pattern at the core, and not by the use of additional circuit elements such as diodes. Of course, it may be possible in some cases to obtain a "sense output" from a selected core large enough to be usable as a word drive current in som type of a word-oriented memory which can use low drive currents, in which case the capability of efficient non-compact addressing is conferred on the word-oriented memory.

The basic design principles and operation of core rope memories are as follows: Each core is threaded by one "set" wire, one "reset"

wire, a number of "inhibit wires, "and a number of "sense wires". The core plus wiring then comprises one "cell," so that the number of memory "cells" is exactly equal to the number of cores. The information in each memory cell is determined by the wiring pattern of the sense wires at that cell. That is, in the bit positions where it is desired that a given cell contain "ones," the corresponding sense wires are threaded <u>through</u> the core; in the bit positions where it is desired that the cell contain "zeroes," the sense wires are detoured <u>around</u> the core. When that core is switched, a much stronger sense signal appears on the lines where a "one" is desired in that word than on the lines where a "zero" is desired, and the respective voltage waveforms for "one" and "zero" are amplified and discriminated by sense amplifier and logic circuits to produce a "one" or "zero" logic level.

The process of selecting a core in the rope to be switched, and hence the method of addressing a cell in the core rope memory, is as follows: Both the true output and the false output of each flip-flop of the address register are amplified, and each amplifier is connected to one inhibit wire; thus 2s inhibit wires are required. Each core is a magnetic NOR circuit; if the inhibit wires threading the core are connected to the $\overline{W}_1$, $\overline{W}_2$, W_3, $\overline{W}_4$, and W_5 flip-flop outputs, the core will produce an output for the minterm condition $W_1 \cdot W_2 \cdot \overline{W}_3 \cdot W_4 \cdot \overline{W}_5$. The logical sum of a set of minterms is obtained by stringing the sense wire through the appropriate set of cores.

It may be seen after some thought that this wiring pattern guarantees that at least one inhibit wire is carrying current through every core <u>except</u> the one which is to be selected. The direction of an inhibit current is such that it tends to reset any core through which it passes, and its magnitude and duration are approximately sufficient to completely switch a core. When the inhibit currents have all been on long enough to stabilize, a current equal to one inhibit current but oppositely directed is passed through the set wire; this "set current" does not affect the magnetization of any unselected core, since it at most cancels the total inhibit current, but it completely switches the selected core and hence "sets" it. After the inhibit currents have been turned off, the selected core is "reset" by a "reset current" in the reset wire; this current has the magnitude and direction of an inhibit current, and hence it does not permanently disturb any core except the selected one, although it passes through every core in the rope. The sense wire outputs can be examined either during the setting phase or the resetting phase of the complete memory cycle, since the selected core is switched during both phases.

There are some variations on the above scheme which are worth comment. One which is obvious, although not always practical, is to use smaller currents which cause only partial switching of the selected core. Another one is to switch all of the cores in the memory except the selected one on the first phase, and then switch it on the second phase: this mode of operation would appear to entail excessive power consumption, but it has never theless found some use. However, a third variation which has been tested in the two core ropes at the Honeywell Systems and Research Division, appears to be extremely worth while; it consists of adding an "odd parity" extra bit to the s-bit address, and connecting this parity bit to a "true" inhibit wire and a "false" inhibit wire just as the other address bits are connected. The parity inhibit wiring is implemented by computing the parity of the address corresponding to each core, and then threading the appropriate one of the two wires through the core and detouring the other one around it. The net effect of this layout is that now a minimum of two, rather than one, inhibit wires are always carrying current through every core except the one which is to be selected. The required amount of inhibit current in each wire can thus be halved, at a considerable saving in drive circuit elements.

Like other types of random-access memory, a core rope memory must be operated more slowly if the number of parallel-readout cells or "words" is increased. Unlike some other types of random-access memory, a core rope memory must also be operated more slowly if the number of bits in each cell (the "word length" of the core rope memory) is increased. The reason that the speed of operation of a core rope memory is sensitive to word length as well as to the number of words is that an increase in either one simply increases the total amount of wire wound through the rope, and the resulting increase in undesirable mutual inductance effects is greater than linear since each new wire interacts with every other wire of the rope. The major disadvantage of core ropes is that it is very difficult to perform all of the required wiring operations automatically, and the manufacturing costs per bit have remained high in comparison to those of other memory types in spite of determined efforts to automate the fabrication process. A modification of the core rope recently developed at the Massachusetts Institute of Technology, Cambridge, Massachusetts, promises to significantly lower these costs. The harness of wires is woven on a loom into a ladder-shaped configuration, and then split transformer cores (in the physical shape of a U with a piece which fits across the U at the top to close the flux path) are placed at intervals along the wire bundle running up one side of the ladder. The transformer cores are not composed of square-loop hysteresis materials, but of linear materials, in contrast to the permalloy

tape-wound and ferrite cores used in conventional core ropes; certain circuit design modifications are therefore required. The resulting memory configuration is called a "braid memory". A modification of this scheme termed the E core memory has also been investigated and has shown considerable promise in terms of speed and economy.

Twistors are a type of read-only memory which have been used by the telephone industry for switchboard "number files," and have been included in a few computers - for instance, the "TIC" (Target Intercept Computer) developed for the Nike-Zeus missile ground control system by Univac, St. Paul, Minnesota. References: "Card-Changeable Memories," J. M. Donnelly, Computer Design, pp. 20-30 June 1964; also "Piggyback Twistor," Computer Design, pp. 20-21 July 1964; also "A High-Speed Card Changeable Permanent Magnet Memory - The Inverted Twister," F. J. Procyk and L. H. Young, IEEE Transactions on Magnetics, March 1965, pages .

3.1.2 Air Coupled Inductive

The air-coupled inductive memory is without doubt the most attractive kind of read-only memory at the present time from a fabrication point of view; there is virtually no simpler memory device in existence. All that the memory itself contains are conductive wires in two orthogonal directions, supported by dielectric substrates so that they cross in close proximity but do not touch. Hence there is nothing to get out of order if the memory is subjected to extremes of radiation, heat, cold, or mechanical shock. The only way that the information can be destroyed is that the memory itself be physically destroyed although the information will certainly become inaccessible to the system in the event of temporary or permanent failure of the semiconductor electronics. A commercially available batch-fabricated memory element in this category is "Permacard." Permacard memories and planes are proprietary custom-item products of Fabri-Tek Incorporated, Edina, Minnesota. The first complete Permacard memory comprised 1024 26-bit words, and featured a .5 microsecond information access time and a 1 microsecond complete read cycle time using a word-oriented configuration of drive and read electronics. A Permacard memory stack consists of alternating sense wire planes and drive wire planes, fabricated by etched copper wiring on flat fiberglass stock. One drive wire plane suffices to store 32 26-bit words in the prototype memory. A drive wire plane and the corresponding sense wire plane are very close together but not in contact; the combined wiring pattern of the two, viewed from

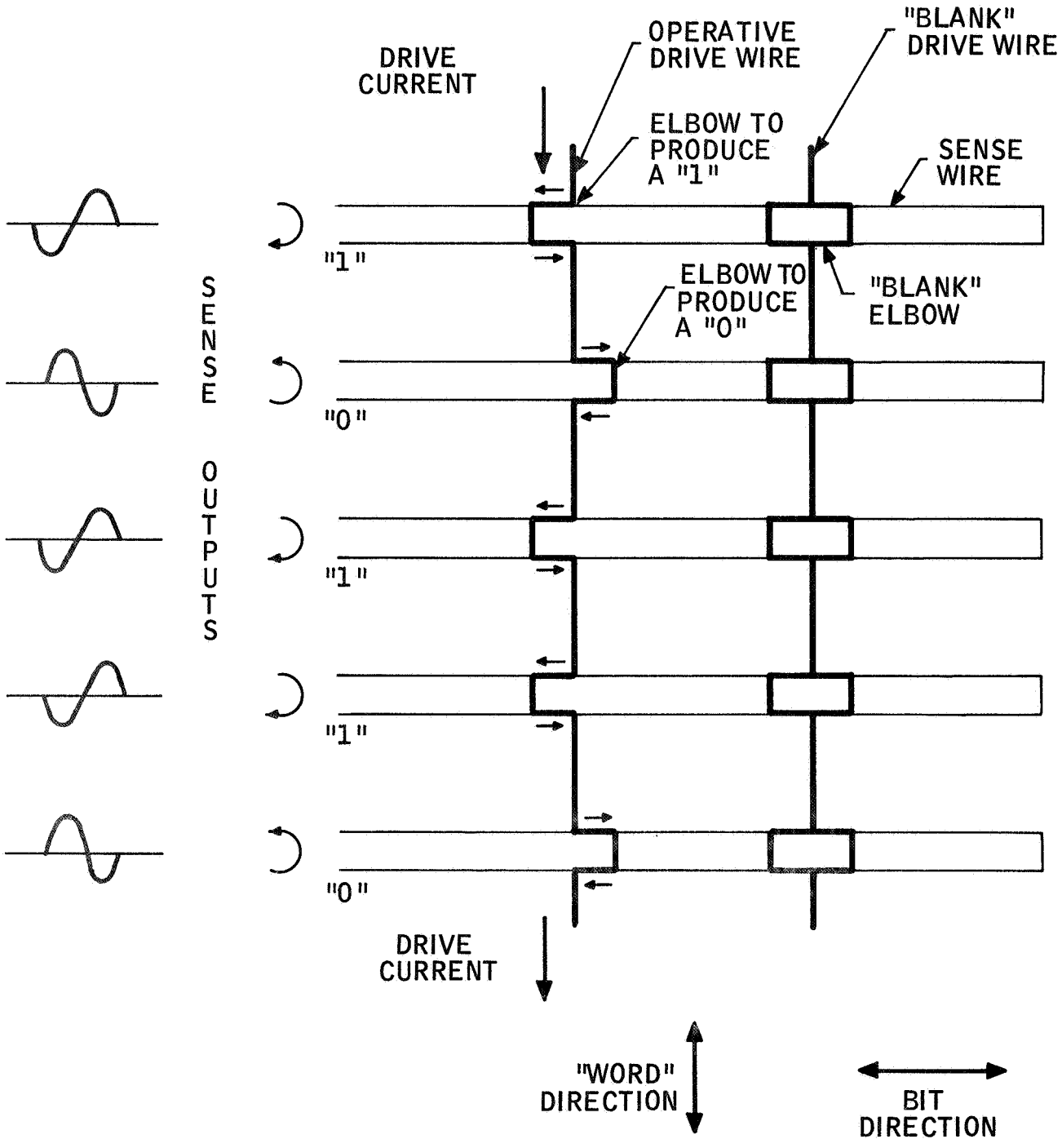


Figure B7. Permacard Schematic

above (assuming transparent planes), is as shown in Figure B7. The drive wire planes are physically removable, and different information may be "stored" in a Permacard memory by an exchange of drive wire planes

The basic operating principle of Permacard (and of other air-coupled inductive memories) is that a drive current is passed down a drive wire, and signals are generated on the sense wires which "intersect" (i. e. cross over in the immediately adjacent sense wire plane) this drive wire by "parasitic" mutual inductance. The polarity of these signals depends on the orientation of the drive wire "elbows," which occur at each intersection and their magnitude is rather small. A pair of intersections is required to store one bit of information, although two pairs could conveniently be used if enhancement of signal strength were desired. The pair is normally chosen so that two adjacent sense wires form a sense wire pair as shown in Figure B7, and a differential sense amplifier is then used. If the same elbow pattern were used for two adjacent drive wires and both were driven simultaneously, a reinforced sense output would result.

On a "blank" drive wire plane, both paths are connected at each drive wire elbow; the information contents of the words corresponding to that drive wire plane may be specified by physically scratching out (with a knife or similar tool) the unwanted path at each elbow, so that the remaining path is oriented properly to produce the output signal for whichever of "one" or "zero" is desired. However, once the pattern of "ones" and "zeroes" to be stored in a particular drive wire plane has been permanently chosen, special printed circuit masks can be made for that plane and it can be produced in any desired quantity, with the information in effect pre-stored.

The geometry of the Fabri-Tek memory has one serious disadvantage; since the drive wires and sense wires are on different substrates, mechanical shock can misalign them so that they no longer register properly. An experimental air-coupled inductive memory plane was designed and fabricated at the Honeywell Aeronautical Division, St. Petersburg, Florida, that is not subject to any such registration problem. This plane is constructed with the drive wires and the sense wires on opposite sides of the same mylar substrate. Another obvious solution to the problem is to deposit the sense wire pattern on a glass or other mechanically stiff substrate, coat the substrate with a dielectric layer by sputtering or other means, and then deposit the drive wire pattern on top of the dielectric layer. In either of these latter cases,

the "changeable card" would now contain both a set of drive wires and a set of sense wires, in contrast to the changeable Permacard drive wire card.

No attempt seems to have yet been made to microminiaturize such a memory, using the smallest practical dimensions for the etched wiring. If this were done, it might very well turn out that an air-coupled inductive memory could be produced with a higher volumetric information density than any other form of magnetic memory. The card-changeability feature could easily be retained under these conditions, although the convenience of changing a single card might have to be sacrificed since a tightly bound stack would have to be disassembled.

There to be sure, certain circuit problems with this type of memory. For one thing, the basic principle of operation of the memory is "parasitic inductance, "and the amount of inductive transfer of energy between the drive line and the sense line has to be just right. If the coupling is too weak, the readout signal will be undetectable; if it is too strong, the signal transferred from one drive line to the set of sense lines in that plane will in turn be transferred to other drive lines and will then in turn be picked up by sense lines again, resulting in background noise problems. Also, it is probably necessary in this type of memory to terminate each drive and sense line in its characteristic impedance in order to prevent "ringing." This would tend to increase the power requirements on the drive circuits somewhat. Nevertheless, the ultimately attainable operating speed of the memory should be that of the drive and sense amplifier circuits; there is virtually no measureable signal delay occuring at the element itself.

Memories of this sort would be suitable for virtually any type of system application contemplated for flat film or plated-wire memories in which there is no necessity to "write." A very plausible application is that of "cold data" memory for instructions and constants in a general-purpose aerospace computer. An associative memory based on air-coupled inductive memory elements might possibly be used in either a bit-slice or local-logic configuration. Their electronic operation is very similar to the NDRO operation of plated-wire memories, except that the magnitude of the drive currents usually employed is generally quite a bit smaller for the air-coupled inductive memories. Since they are highly compatible with plated wire memories, it might be possible to develop an associative memory having very long words, in which three quarters of the word length was fixed air-coupled inductive memory and one quarter of it was electrically-alterable plated-wire memory.

If the electronic problems of readout in these memories can be effectively solved, extremely high information density and readout cycle times down to 100 nanoseconds should be possible. This type of memory therefore deserves serious consideration for the full range of read-only memory applications. There are some other types of "air-coupled inductive memories" which differ considerably from the basic Permacard philosophy. One of these that uses actual wires and not printed writing is the "unifluxor" memory, a forerunner of Permacard, which was developed by Univac, St. Paul, Minnesota. (See "Unifluxor: a Permanent Memory Element," A. M Renard and W. J. Neuman, pp. 91-96, Proceedings of the Western Joint Computer Conference, May 3-5, 1960..) Some other schemes based on mutual inductance are described on pp. 24-28 of a recent technical magazine article ("Card-Changeable Memories," J. M. Donnelly, Computer Design, pp. 20-30, June 1964.)

A memory system which is somewhat different from the others just described has been developed by Sylvania Electronic Systems, Waltham, Massachusetts, it is a read-only card-changeable memory, based on long, thin, air-core solenoids. The characteristics of this memory are such that, to specify its information contents, a desk of special cards is punched out on any standard computer-controlled tabulating card punch. These special cards have holes pre-punched for the solenoids to pass through; after being punched, the card deck is collected into a stack and the solenoids are inserted. Although there must be one solenoid-card intersection per bit in the memory, if the array is to be used for other purposes than those of a binary information memory it is possible to use solenoids of varying "weights" and to generate a sense output having up to 15 distinct levels. Sylvania seems to have intended the solenoid array primarily for usage in this multi-level mode, which is economic in the type of pattern recognition problems which originally motivated the development. Two references on Sylvania's memory, both in the February 1964 issure of the IEEE Transactions on Electronic Computers, are: "The Solenoid Array, A New Computer Element," G. G. Pick, S. B. Gray, and D. B. Brick, pp. 27-35; and "Microsecond Word Recognition System," D. B. Brick and G. G. Pick, correspondence, pp. 57-59.

3.2 Capacity Coupled

For virtually every magnetic field effect, there is an analogous electric field effect. This analogy must be used with caution, but it is generally helpful in reasoning about computer memory elements. Capactitivity coupled read-only memories are suggested by the analogy; they exist, but have not been investigated to the same extent inductively coupled read-only memories.

3.2.1 Ferroelectric or Ferrielectric Coupled Capacitive

Any of the various schemes suggested in section 2.5, for memories composed of ferroelectric or ferrielectric elements, could be modified so that a 1 state was represented by the presence of such an element at the intersection of a drive wire and a sense wire, and a O state was represented by the absence of any such element. The drive-wire-sense-wire coupling would be very strong in the first instance, and very weak in the second instance. This type of memory might have the advantage, over the analogous magnetic memory, of much lower power consumtion. It does not appear that there has yet been any work done on this type of memory although perhaps there should be.

3.2.2 Air-Coupled Capacitive

Several schemes have been developed for using batch-fabricated thin film capacitors for read-only memories. Here again, a 1 state is indicated by the presence and a 0 state by the absense of a capacitor which couples a drive line to a sense line. In this instance, the capacitor merely consists of thin metallic layers on opposite sides of a dielectric layer.

A batch-fabricated thin-film capacitor may be constructed as follows: two leads on one side of the dielectric layer are terminated in pads which almost, but not quite, meet. A disk-shaped layer of conductor is deposited on the opposite side of the dielectric layer, so that it overlaps both pads. The resulting structure is two (approximately equal) capacitors in series, and behaves exactly like one capacitor of half the capacitance.

As in the case of the air-coupled inductive memory, there are circuit problems with this type of memory which require further study. Also, the basic physical structure duplicated at each coupled intersection of a drive line and a sense line is somewhat more complicated than that analogous structure for an inductive memory; some versions of the capacitor approach would definitely require more manufacturing steps. On the other hand the average power level required to operate the memory might be somewhat lower than that required to operate an inductive memory.

A reference concerning air-coupled capacitive (and other) read-only memories is "Card-Changeable Memories," J. M. Donnelly, Computer Design, June 1964, pages 20-30.

3.3 Diode Coupled

At the intersection of a drive line and a sense line, the presence of a diode may be used to indicate a 1 state and the absence of a diode may be used to indicate a 0 state. Potentially, such a diode array would seem

promising as an ultra-high-speed read-only memory; practically its speed is adversely affected quite severely by any increase in size. At present, the fabrication problems involved in making this sort of memory with discrete diodes are rather serious, but batch-fabricated integrated-circuit techniques can probably be worked out.

A diode matrix memory comprising 2,048 18-bit words has been developed by the Honeywell Aeronautical Division, Minneapolis, Minnesota. It has a readout access time of 28 microseconds which is very slow; however, such a memory could be many times as fast if it were not as large. Autonetics, Downey, California, has developed a batch-fabricated sapphire diode memory. Univac, Blue Bell, Pennsylvania, has developed a thin-film diode logic array; see "Batch Process Thin Film Diode Logic Array and Their Evaluation, J. S. Cubert and J. J. Murphy, Proceedings of the National Symposium on the Impact of Batch Fabrication on Future Computers, 6-8 April 1965, pages 53-66.

A batch-fabricated diode array might be useful as a high-speed fixed-incormation associative memory where the number of words is quite small, but it is not easy to see any really wide field of application for this type of memory.

3.4 Serial Access

"Strain-wave" memory devices are described in subsection 4.2.1. These are based on phenomena controlled by the propogation of an acoustic wave along a fiber, in particular the change in shape of a ferromagnetic or ferroelectric hystersis loop in material bonded to the fiber as the strain wave passes. If a strain-wave memory is fabricated in the normal way, and then the hysteresis material is physically removed for each points which is to correspond to a 0 state and allowed to remain at each points which is to correspond to a 1 state, the same type of NDRO operation normally used with such memories (see subection 4.2.1) can be used to read out permanently stored information. Apparently no memory of this type has ever been built, but it might be of interest in any design situation where both the reliability of a read-only memory and the simplicity of serial operation were desired. Such a specification might very well arise in the design of small-scale general-purpose computers or read-only associative memories for extended space missions.

4.0 CYCLIC-ACCESS MEDIA

The term cyclic-access media signifies "a memory circuit device storing many bits, which has the property that readout of information must be carried out in a certain predetermined order rather than on a random-access basis." The majority of such devices have the property that, if readout proceeds far enough in one direction, the information read out begins to repeat the sequence of bits already obtained, hence the term "cyclic." There are a few devices, such as magnetic tape units, which have been placed in this category even though the term "cyclic" is not entirely appropriate for them; however, there does not seem to be any other term which is equally suggestive and more appropriate.

It is expedient to partition the category of cyclic-access media according to their timing properties, rather than according to the materials comprising them or the physical principles involved in their operation. There are two basic timing characteristics of a cyclic-access memory: one has to do with the initiation of an "information transfer: (a read or a write) operation, and the other has to do with the clocking of the read or write operation once it commences. Each property may be "synchronous" or "asynchronous." Initiation must be "synchronous" in the case of media in which the information is traveling by the reading electronics at a constant speed at all times; but for other types of media it may be specifiable at an arbitrary time, and hence may be "asychronous." If the transfer of information must be done at a constant clock frequency, clocking is said to be "synchronous"; if one step in the information transfer operation may be carried out whenever an arbitrarily times pulse is received, clocking is said to be "asychronous".

A common and useful synonym for the term "asychronous" is the term "demand." An asychronous device is one which may be operated "on demand," where the "demand" may be made by some different piece of equipment.

Asychronous (not necessarily cyclic-access) devices are used throughout data processing equipment for information "buffering" purposes, where an information transfer rate governed by some external piece of equipment must be matched by the control processor. There is reason to believe that data acquisition systems, involving the sampling and control of many external devices by a control processor, will represent a attractive area for future associative computers.

4.1 Synchronous Initiation

Sychronous initiation cyclic-access media are those in which either the information, or the physical memory medium itself are constantly in motion; and the reading electronics determine where they are reading by noting the time at which they are reading with respect to a reference time, and similarly for writing. The classic example of a cyclic-access synchronous initiation medium is the magnetic drum. There are also many geometrical variations of the drum, which are called by other names. There is also a second major class of cyclic access media devices called delay lines, which as memories are logically similar to drums, but as physical devices are very dissimilar.

4.1.1 Synchronous Read-Write

Most of the storage devices which were used in early central processor memories for relatively small-scale computer systems fall into this category. It includes at least five types of devices of some

commercial importance: magnetic drums, rigid magnetic discs, flexible magnetic discs, torsional mode nickel delay lines, and glass delay lines. Three other with occasional application, or some promise of application to computers are: longitudinal mode nickel delay lines, quartz delay lines, and cryogenic delay lines. The properties of most of these are quite well known. The drums and discs are rotary magnetic media; information is read and recorded using by magnetic "heads," mounted so they are very close to the magnetic recording surface as it rotates past. The same head is normally capable of performing both reading and writing although the associated reading and writing electronics are different. Each head consists effectively of a small horseshoe magnet with many turns of wire. Reading is performed non-destructively, by sensing the changes in flux in the magnetic head windings as a particular magnetized spot on the medium passes by close to the head. Writing is performed by sending a drive current through the head windings, at the time that the proper spot on magnetic medium is beneath the head.

Cylindrical drums were the first widespread form of small computer memory. They are still produced and still sell at some volume, but have now been relegated to the role of "backup" storage instead of being used for central processor memories. The drums and disks made in recent years are very much better than the drums which were used in early central processors; information packing densities along the recording track have been increased from 100 bits per inch or less to as much as 800 or 1000 bits per inch. This increase in recording density has come about through great improvement in the quality of magnetic coatings, in the control of tolerances in manufacturing both the drum itself and its housing, and in magnetic read/write head design. The mechanical problem of getting the heads close enough to the surface of the rotating drum was solved by the development of various aerodynamically floated head configurations, and this development was also a factor in the increasing use of the disk geometry. To be sure, the disk geometry has one immediate disadvantage over the drum geometry, in that all tracks around the disk are not of equal length and hence the recording density must be higher in some tracks than in others; whereas, in the drum all tracks are around the circumference and are the same length, and recording densities are uniform. However, disks are easier to machine than drums. Moreover, a disk is more physically compact than a drum, and careful systems mechanical design of disk memories has resulted in much greater information density per cubic foot than is feasible with drums. Bryant Computer Products of Walled Lake, Michigan supplies disk file modules with capacities ranging up to a billion bits; they are comparable in physical size to some large drums with a thirtieth of the storage capacity.

Not all disk memories are intended for mass memory applications; some have been engineered to serve as small central processor or buffer memories. The most famous of these is probably

for the one designed some years ago by Autonetics, Downer, California, for the Recomp series of Computers. The disk head arrangement in these machines was aerodynamically designed on the basis of boundary layer theory to "float" the disk when it achieved operating speed, so that the heads would be extremely close to the disk but not close enough to touch; the disk was free to move along its shaft to a limited extent. The same problems is solved in a different manner in disk mass memories; they typically have the heads mounted at the end of some sort of aerodynamic "shoe" which automatically moves up and down slightly to compensate for minor irregularities in the surface of the disk, and is maintained at a proper distance from the surface of the disk by boundary layer effects. A radical approach has been taken at Laboratory for Electronics, Inc., Boston, Massachusetts; LFE has developed a family of flexible mylar disks, which distort themselves until they attain the precise shape of the head assembly when rotating at full speed.

Delay lines comprise a class of devices having logical similarities to rotary magnetic media, but no physical resemblance. Information is propagated along a delay line in the form of some type of wave instead of having the medium itself move mechanically with respect to fixed-position/write heads. In most cases the information is stored in the form of acoustical torsion or compression/rarefaction pulses, which are started at one end of the line by an electrostrictive transducer, propagated along the line, and are picked up at the other end of the line by a piezoelectric transducer which in turn feeds them back through an amplifier to the electrostrictive transducer. The first really successful delay line memory was the torsional-mode nickel delay line, which was first used in several computers built by Ferranti Ltd. of Manchester, England, and more recently in the Packard Bell 250 computer built by the former Packard Bell Computer Company as now called Raytheon Computer. Torsional mode nickel delay lines have been markedly improved; such a line exists with a delay of 3-milliseconds which is capable of being operated at a 2-megacycle bit rate. This means that the line has a total information capacity of 6,000 bits. Somewhat higher speeds may soon be practical.

Glass delay lines, on the other hand, can be operated at bit rates of up to 70 megacycles on a laboratory basis, although reliable operation at this rate in practical digital memories is not contemplated. These devices use a compression/rarefaction pulse, and are designed so that many internal wave front reflections occur in a glass prism, in order to obtain long delay times. Glass delay lines are available from the Corning Glass Company, Corning, New York. They are capable of being operated over a substantial range of temperatures. Their information capacity ranges up to over 40,000 bits per line, which would correspond to a delay of 10 milliseconds and an operating bit rate of 40 megacycles.

Quartz and fuzed silica delay lines are available from Anderson Laboratories, Inc., Connecticut. These devices are apparently

apparently operable at a somewhat higher information rate than glass lines but their characteristics are less stable with changes in temperature. The principles of operation of these lines are essentially the same as those of Corning glass delay lines.

Nickel longitudinal (compression-rarefaction) mode delay lines are also available, but they provide delays of at most a few microseconds. Although they are operable at somewhat higher information rates than nickel torsional mode delay lines, their information storage capacities are so low that they can only be used as one-word registers in computer systems, rather than main memories as is quite feasible with nickel torsional mode and glass delay lines.

"Cryogenic delay lines" have been developed by the Martin Division of Martin-Marietta Corporation, Baltimore, Maryland; these are purely electrical devices, with no acoustic principle. A cryogenic delay line is simply a length of coaxial cable kept in a cryogenic insulating container called a "dewar." Electrical pulses inserted at one end will appear at the other end about a microsecond later without excessive attenuation; the resolution of the device is sufficiently sharp that operation at information rates of up to 125 megacycles is claimed. Unfortunately, this rate considerably exceeds the state-of-the-art in low-cost digital circuits. Cryogenic delay lines may find some use in the future as serial one-word registers associated with serial arithmetic units; their information capacities will have to be increased before they will be suitable for many other applications.

All of these various rotary and delay-line media could be used in associative memories organized in such a way that a particular storage line (that is, a delay line or a track on a drum or disk associated with one read-record head) would correspond to a given bit position in all the words of the memory. Bit slices would then be available at the output in sequence at a fixed synchronous rate. External logic would be required to sense matches and mismatches and to do arithmetic. The Honeywell Aeronautical Division, St. Petersburg, Florida, has investigated glass delay line associative memories organized in this manner.

4.1.2 Asychronous Read-Write

This category is included for logical completeness, but there are no such devices in existence at present. To fall into this category, a device would have the property that the initiation time for an information transfer operation by a storage device itself, because of some periodic synchronization characteristic; but, once the information transfer process had commenced, each step in the process would be performed "on demand". It is difficult to imagine what sort of physical memory device or information storage principle would have these characteristics.

4. 2 Asychronous Initiation

Devices in this category are "at rest," in some particular sense, when reading or writing is not actually taking place. In a delay line, information is constantly cycling; in drum and disk memories, the medium is cycling constantly. In elements of this category however, information is obtained only upon demand.

4. 2. 1 Synchronous Read-Write

There are three classes of elements presently in existence in this category; strip tape memories, strain-wave memories, domain-wall addressed memories. The latter two are of considerable interest for associative memory applications.

A strip tape memory contains many magnetic tape cartridges, each of which contains a relatively short strip of tape. Upon receipt of a command to read or write, the strip tape memory electronically selects one particular cartridge. The tape is read or written much as in a conventional tape transport, except that there are no take-up reels provided; as the tape is moved, it is simply dumped into a bin, and upon the conclusion of the read or write it operation is repositioned so that the front end of the strip is again ready for access. In some such memories the tape forms a continuous loop.

Strip tape memories have been experimented with for many years by computer system manufacturers. Hughes Aircraft Company, Culver City, California attempted to develop one almost ten years ago, but gave up because the tape then available would take a permanent "crinkle" over night after sitting in the bin. Somewhat later, strip tape memories were made available with the by Librascope for use with the RPC 4000 computer. Most recently, Potter Instrument Corporation, Plainview, Long Island has put on the market a large-scale, up-to-date strip tape memory under the trademark RAM (Random Access Memory).

The next class of devices, strain-wave memories, is of serious interest for avionic associative memories. Strain-wave memories offer the low cost per bit and modularity features of exiting types of delay lines (see subsection 4. 2. 2), without either of their serious drawbacks that the information contained is constantly recirculating at a fixed rate, and that the information is lost in the event of power failure. Strain wave memories have been investigated by General Dynamics/Electronics, Rochester, New York; RCA Research Laboratories, Princeton, New Jersey; and Sylvania Electronic Systems, Waltham, Massachusetts. Most of this work has been done under contract to U. S. Electronics Command, Fort Monmouth, New Jersey in connection with the "BORAM" (Block Oriented Random Access Memory) development program, whose purpose is to come up with a device having the systems applicability of a magnetic tape unit but with no moving parts. General Dynamic also had a NASA contract in this area.

RCA has pursued only magnetic thin-film strain-wave memories, whereas General Dynamics has worked on ferroelectric strain-wave memories as well.

At the present stage of development of strain wave memories, it is not possible to speak with much confidence about their ultimate capabilities. Their basic principle of operation is as follows: The hysteresis loop of "square-loop" ferromagnetic and ferroelectric materials actually varies in shape considerably (except for "zero-magnetostrictive" magnetic materials), depending on whether the material is unstressed or whether it is under compression or rarefaction. In general, the application of a stress causes the material to switch with lower applied field strength. Hence if a strain pulse is propogated down an elastic fiber coated with square-loop memory material, and a polarizing field of just the right magnitude is then applied across the fiber, switching will occur in the vicinity of the strain pulse but nowhere else; that is, the material is switched by the coincidence of a field and a strain pulse. Non destructive readout can be performed simply by passing a strain pulse down the fiber, and then examining the output signal, which in the magnetic case is obtained from a linking conductive plating around the fiber under the magnetic plating layer, and in the ferroelectric case is obtained between two contacts strips on opposite sides of the ferroelectric-coated fiber strip; the hysteresis loop changes caused by the passing of the strain-pulse produce a signal in each case.

Since the transducer which originates the strain pulse can be activated at an arbitrary time, initiation of an information transfer is asychronous. Once this operation has been initiated, though, the rate of information transfer is governed by the speed of propagation of the strain pulse in the fiber, which is a physical constant (the "speed of sound"), and by the information packing density along the fiber, whose upper limit is determined by the acoustic resolution properties of the fiber material; hence, clocking is synchronous.

The properties of strain-wave memories are very appealing for an avionic arithmetic-search associative memory; the same simple serial organization could be used which has been proposed using glass delay lines, but now the stored information is non-volatile; the memory is inactive (and not consuming power) except when a search operation is requested, and the operation may be requested to begin at an arbitrary time instead of only at the beginning of a word recirculation. Moreover, the memory medium is continuous, batch-fabricated, and hence potentially very low in cost.

General Dynamics successfully operated a laboratory prototype magnetic strain wave memory at 330 kilocycles, and tested elements at double that rate. They estimate that ferroelectric strain-wave memories may achieve information reates of up to 12 megacycles, although such rates have apparently not been demonstrated. RCA's program was earlier, and by their own account much less successful. Both RCA and General Dynamics have taken approaches based on

coating individual fiber; Sylvania, , on the other hand, apparently has an approach which may be more readily suited to batch fabrication in which parallel piezomagnetic film strips are deposited on a flat substrate capable of supporting a strain wave.

The metallic compositions required for magnetic thin-film strain-wave memories are markedly different from those used in plated-wire memories, since in the former case hysteresis-loop changes as a result of strain are a desirable property rather than an evil. General Dynamics has been able to construct a number of short magnetic strain-wave lines which performed well in a laboratory prototype memory.

The ferroelectric strain-wave memory requires material which are similar to compositions already required for other purposes; for example, lead zirconate-lead titantate. This material is used in electroacoustic transducers and is marketed by Honeywell, Golden Valley, Minnesota, under the trade name PZ-PT, and by Clevite Corporation, Cleveland, Ohio, under the trade name PZT. Many different PZ-PT compositions are possible by varying the proportions of zirconate and titanate and by adding various impurities. Strain wave devices would be used in associative memories in essentially the same manner as synchronous initiation devices such as drums and delay lines. There would be one such device per bit position in a word, and hence bit slices would be obtained from the output sense amplifier in sequence. Matching and arithmetic operations would be performed by per-word semiconductor logic. If a separate transducer were used for generating the strain pulse for each word line, it would be possible to delay activation of one transducer in order to get different fields in two words aligned for comparison, addition, or subtraction; hence, such a memory could perform multiplication and division, and other operations requiring shifting. It would be natural to store information by filling an entire word line at once; the entire line would be filled just like an ordinary acoustic (nickel, glass, etc.) delay line, with strain pulses, and then the magnetic or electric field would be applied along the entire line all at once to in effect "freeze" the information, present in the form of compressions and rarefactions, at its instantaneous position into a non-volatile magnetic or electric polarization of the storage medium. After this storage operation has been performed, readout can be non-destructive, and in principle it can occur any number of times without degradation of the stored information.

The ability to select a chosen field of a word line implies a counter, which may be synchronized to the strain-wave itself in such a way that the counter contains a count equivalent to the number of bit positions that a strain wave has propagated down each line in which it was initiated. One way to build such a counter, which would have the advantage that it would automatically be temperature-compensated with respect to the word lines comprising the memory, would be to store bit position counts as parallel data in several lines

in parallel. For instance, if one had 512 positions in a word line, there would be nine counter lines which would have counts stored in them in a word-slice configuration, such that as simultaneously initiated strain pulses traveled down all nine lines together, all counts from 0 to 511 would successively appear on the sense amplifiers of these nine lines. Magnetic strain-wave lines have been operated for test purposes at from -20 degrees to 85 degrees centigrade, and hence there is reason to believe that strain-wave memories can withstand aerospace environments.

Another type of memory making use of strain waves might be implemented by interrogating a ferrotron memory with a strain wave line controlling a continuous light source. It has been shown by General Telephone and Electronics Research Laboratories, Bayside Lone Island, that a sandwich consisting of an electroluminescent layer, a non-linear resistance layer, and a piezoelectric layer has the property that the passage of a strain wave along the piezoelectric layer may be accompanied by luminescence of the electroluminescent material at the point of maximum strain amplitude, and nowhere else. General Telephone has used this effect in flat plates which propagate strain wave fronts in two orthogonal directions; the intersection of these two wave fronts traces out a line diagonally across the flat plate. This effect, with proper timing can become a means of generating faster lines for a television display; GT&E made and operated a flat television screen of this type with a five-inch picture face. If this sort of device were used to generate interrogation light strips to select rows or columns of information on a ferrotron plate, the ferrotron memory would then have the timing properties of a strain wave memory.

Burroughs in presently investigating a thin-film memory with logical properties similar to those to strain wave memories, in which the propagating wave is a domain wall in a magnetic film layer subject to a d-c bias. The domain wall is started at one end of the film layer, and the bias field causes it to propagate down the film toward the other end. As it passes each particular area the flux disturbances which it causes can produce NDRO output signals from information storage elements in an adjacent coupled film layer.

References on strain-wave memories include various contract reports available from ASTIA and other sources, as follows:

Digital Computer Peripheral Memory, Reports 1, 2, 3, 4, Contract No. DA36-039-AMC-03288(3), staff of RCA Laboratories, Princeton, New Jersey. Dates are 30 September 1963, 31 December 1963, 31 March 1964, and 30 June 1964 respectively. Report 4 is the final report; its ASTIA Document number is 449506.

Electroacoustic Digital Computer Peripheral Memory, Reports 1, 2, 3, 4, Contract No. DA28-043 AMC-00267(E), staff of General Dynamics/Electronics, Rochester, New York. Dates are 30 September 1964, 31 December 1964, 31 March 1965, and 31 June 1965 respectively. ASTIA Document numbers for the first three reports are 454599, 461843, and 466687.

Research on a Ferroacoustic Information Storage System, same contract number and contractor as preceding report, NASA Contractor Report CR-45, June 1964. "Ultrasonic Approach to Data Storage," J. W. Gratian and R. W. Freytag, (General Dynamics/Electronics), Electronics, 4 May 1964.

4.2.2 Asychronous Read-Write

The elements in this category are customarily called "shift registers." Shift registers may be produced entirely out of semiconductors and passive elements, and it is now possible to buy integrated-circuit shift registers of various descriptions, but these will not be discussed in this subsection. The types of shift registers of interest for associative memory applications are those which are based on a batch-fabricated hysteresis element. One type which is now available commercially is the magnetic domain wall motion shift register which has some attractive properties; it stores information in a relatively compact form, it shifts on demand in either direction with equal ease, it retains information if power is lost, and its reliability appears to be quite good.

The first company to have offered a magnetic domain wall shift register commercially apparently was Servomechanisms Inc., Goleta (Santa Barbara), California. The original model, as advertised in September 1961, stored 23 bits and could be operated at any frequency from DC to 150 kilocycles. Kent Broadbent, then at Hughes Aircraft, Culver City, California, began working on this type of shift register many years ago and obtained a patent (assigned to Hughes) in 1959. The extent of connection between his work and that of Servomechanism is uncertain, but the type of shift register developed in both cases; appears to be basically the same. Broadbent subsequently moved to American Systems Inc., and the new company began offering domain-wall motion shift registers commercially. Although American Systems was later dissolved, Broadbent maintained his thin film operation in business for another period of time under the name of Broadbent Laboratories. Most recently Broadbent Laboratories was acquired by Interstate Electronics Corporation, Los Angeles, California, which has energetically marketed the shift register. Several generations of the same basic shift register have thus been offered for sale by American Systems, Broadbent Laboratories and now Interstate. Another Hughes division in Fullerton California has resumed the investigation within Hughes, as will be discussed shortly.

A third basically separate line of development has occurred at Bell Laboratories, Murray Hill, New Jersey, where domain-wall motion shift registers have been developed by U. F. Gianola and J. T. Sibilia. Unlike the Servomechanism's and Broadbent shift registers, which use a flat thin film on a flat substrate, the Bell Laboratories design uses a permalloy wire wrapped around a cylindrical tubular substrate. The permalloy composition used is relatively close to the usual zero-magnetostrictive composition except that some of the iron has been replaced with niobium and silver so that the final alloy is 79.4 percent nickel, 16.6 percent iron, 3 percent niobium, and 1 percent silver. Bell intends to manufacture this device at their facility at Allentown, Pennsylvania. A similar type of memory device is now being investigated at Hughes Aircraft, General Systems Division, Fullerton, California, as a possible BORAM (see subsection 4.2.1); Hughes keeps the wire stressed to 75 percent of its yield strength (probably to determine a particular shape of hysteresis loop).

The operation of the domain wall motion shift register is based on the fact that, once a domain has been created in a magnetic material that is required to create a new one. Hence, by the application of magnetic fields spaced at well-chosen intervals, the domain once formed can be moved along through the magnetic material without much change in its shape. For conservation operation, the field required to create a domain should be at least twice as large as that required to move it through the material; this is approximately the ratio obtained in the flat thin film devices. In the Bell Laboratories permalloy wire device, the ratio is closer to ten to one.

The operation of one of these shift registers typically requires a four cycle drive. A drive wire that is used to supply one of the fields is run back-and-forth across the surface of the shift register or lengthwise back and forth on the cylinder. A second, similarly configured, drive wire alternates with the first one; the second wire deposited on the other side of a thin substrate, or underneath a thin dielectric layer such that it is electrically isolated from the first wire. At the beginning of a cycle, the first wire receives a pulse of one polarity (assume "positive" for convenience); then the second wire receives a positive pulse, the first wire receives a negative pulse, and the second wire receives a negative pulse. When this pulse sequence is completed, a bit stored as a domain has been advanced over one complete wire loop; the reason for the use of negative pulses is that the drive wires reverse direction because they are laid out in a back-and-forth pattern, and the same sense of field is always desired at the magnetic material as long as the stored information is still being propagated in the same direction.

Bell Laboratories has built a test model which stores 2,000 bits and operates at 40 kilocycles. They plan to increase this to 150

kilocycles, and feel that 250 kilocycles can ultimately be achieved in their type of device. Interstate claims that their device has been operated at 4 megacycles in the laboratories; they have delivered custom units operating at two megacycles, and their standard commercial version operates at 1 megacycle.

Domain wall motion shift registers undoubtedly will remain more expensive per bit than other types of cyclic-access memories. For one thing, their manufacture requires the fabrication of four crossings (of a drive wire with thin film strip or magnetic wire) per bit of information storage capacity; thus these devices are only to a limited extent "batch-fabricated." There are some associative memory applications where their properties are would be very suitable, particularly those in which buffering of serially organized information is required. As associative memory composed of domain wall shift registers like this would be organized with many of them in parallel, propagating domains in the "word direction," with a bit slice read out from the ends of all of the registers in parallel, and in a cyclic arrangement in which information read off the end was immediately sent back to the beginning. One feature of this type of memory is that it would be relatively easy to attach a number of output transducers, at various points along each register, instead of just having one at the beginning and one at the end as is the normal procedure with other types of cyclic-access memories. It would thus be possible, in an associative memory composed of shift register word lines, to access several fields of the same word simultaneously. Moreover it would also be possible to cause the information to shift "back" as well as "forward" one bit position at a time. There may be some data sampling applications where information is naturally available a bit slice at a time, at a rate synchronized by some external device; in such cases an asychronous memory which can store information as it is received, without the need for generation of a random-access address, would be very suitable.

The basic paper on magnetic domain wall shift registers is "A Thin Magnetic Film Shift Register," K. D. Broadbent, IRE Transactions on Electronic Computers, pages 321-323, September 1960.

The possibility would also seem to exist of ferroelectric domain wall shift registers, but it is not clear that the proper energy relationships can be satisfied with present ferroelectric materials. The amount of energy needed, to move a domain wall in a ferroelectric material, is approximately equal to that required to form the wall in the first place. One source, Solid State Physics, A. J. Decker (Prentic-Hall Inc., Englewood Cliffs, New Jersey, 1957) states that the domain walls of a ferroelectric material have a thickness of the order of several hundred lattices distances, and the energy required to move the wall one lattice distance is inversely related to the wall thickness. Ferroelectric shift registers have been built, but they do not use the domain wall motion phenomenon, and they require many

semiconductor elements because the information is actually read from one element and written into the next. This scheme is described in "A New Type of Ferroelectric Shift Register," J. R. Anderson, IRE Transactions on Electronic Computers, December 1956, pages 184-191. As presented there this type of shift register is quite slow; the highest speed operation contemplated seems to have been 10 kilocycles, but this limit was based on the ferroelectric material and the thin film technology of 10 years ago, and great advances have been made in both areas.

Before the domain wall motion thin film shift register made its debut, the principal type of magnetic shift register available was composed of magnetic cores. One form was constructed from five-hole "MAD" magnetic cores operating in a complex transfluxor mode; these registers are made by AMP Inc., Harrisburg, Pennsylvania. Their operating information rates go up to about 10 kilocycles. Another older type of magnetic shift register is based on the use of ordinary toroidal cores and semiconductors; these are made by Di/An Controls, Inc., Boston, Massachusetts.